The Science of Educational Research

GEORGE J. MOULY
University of Miami

EURASIA PUBLISHING HOUSE (Pvt.) Ltd
RAM NAGAR, NEW DELHI-1

6.10.93
7460

Sole Distributors

S. CHAND & CO.

Ram Nagar	—	New Delhi
Fountain	—	Delhi
Mai Hiran Gate	—	Jullundur
Hazrat Ganj	—	Lucknow
Lamington Road	—	Bombay

This book has been published with the assistance of the joint Indian-American Standard Works Programe.

First Indian Reprint : 1964

Published in India by Eurasia Publishing House (Pvt.) Ltd., New Delhi, by arrangement with The American Book Co., New York.

Copyright © 1963 by

AMERICAN BOOK COMPANY

Mouly: *The Science of Educational Research.*

All rights reserved. No part of this book protected by the above copyright may be reproduced in any form without the written permission of the publishers.

Offset by Eurasia Offset Printers, Ram Nagar, New Delhi-1

STATE INSTITUTE OF EDUCATION
BANIPUR

Preface

THIS text is written for all those who are interested in educational research. It is oriented specifically to three major groups: 1. graduate students who are working toward the fulfillment of their research requirements; 2. graduate students or advanced undergraduates who need to be aware of the methods and procedures of research; and 3. teachers and administrators who are interested in the solution of their everyday problems.

The text provides a background both for the *producer* of research concerned with the promotion of education as a science, and for the *consumer* of research interested in the interpretation and application of research findings. Although no text can give an adequate coverage of a field as broad as educational research, this book does provide an orientation to the nature of research, the procedures by which it is conducted, and the crucial role it can play in the advancement of education as a science.

The author has attempted to make the text practical without sacrificing the emphasis on theory and science which gives research unity and meaning. Space limitations have precluded the extensive treatment of certain topics. An orientation to the more pertinent aspects of educational research is provided, and the student is referred to sources more specifically designed to deal with some of the more advanced or comprehensive topics.

The area of statistical analysis is worthy of special mention in this connection in view of its significance as a tool of research. Although it is not the purpose of this text to provide training in the mechanics of computing common statistical measures, the

student needs to be cognizant of their existence and their significance. Any serious student of research must sooner or later become familiar with these procedures and develop a certain proficiency in their use. Since adequate treatment of these procedures is readily available in sources where they more correctly belong, they are not discussed in detail in the present text. The field of tests and measurements, another crucial area, falls in the same category.

THE SCIENCE OF EDUCATIONAL RESEARCH is concerned with research methods, and it will leave the presentation of actual research studies to the instructor, the *Encyclopedia of Educational Research,* the *Review of Educational Research,* and other professional journals.

This text is not easy; it is not meant to be. Science and research are far too complicated and exacting to be easy. On the contrary, the text is based on a philosophy that the upgrading of education, triggered by Sputnik, but inevitable nonetheless, must be reflected not only in more adequate "education" for school children, but also more thorough training for their teachers. There is need for a greater orientation of education toward research as the key determinant of its relative success. It is high time we subject the education of our country's children to the same scientific treatment we give to the material aspects of their everyday life. It is time that educators assume their rightful place among their fellow scientists.

Contents

PART I: SCIENCE AND THE SCIENTIFIC METHOD

I feel we are on the brink of an era of expansion of knowledge about ourselves and our surroundings that is beyond description or comprehension at this time.

LT. COL. JOHN GLENN

1 The World of Science

MODERN SCIENTIFIC ADVANCES

Material Progress

We live in a world of fantastic scientific achievements ranging from those which contribute to the maximum welfare and pleasure of man to those which are capable of his complete annihilation. We have conquered time and distance, the sea and the sky; our atomic submarines can stay an unlimited time and cover an unlimited distance under water; our aircrafts can travel faster than sound. We have placed satellites and astronauts in orbit, and we are on the verge of interplanetary travel. We have bombs and missiles capable of the almost instantaneous destruction of whole cities. Even the tales of H. G. Wells are rapidly assuming an air of plausibility, if not reality.

Medical research has made spectacular discoveries in the prevention and cure of certain diseases. It has effected such a drastic reduction in mortality rates in the underdeveloped nations of the world that we are threatened by a population bomb of even greater potential danger than its atomic counterpart.

Over half the products of modern industry were unknown at the beginning of this century, and patents for new and improved products are being issued at the rate of one

every ten minutes. We operate push-button factories while electronic computers process in seconds data that no more than fifty years ago would have taken a lifetime of patience and toil. What's more, the pace is rapidly increasing. What we are seeing now is only the beginning!

Generally our achievements have been a blessing, but some have created problems of readjustment. For example, we have eliminated the universal problem of starvation, but, at least in America, we have created the opposite difficulty. Automation has increased our power of production beyond our ability to consume so that we are now faced with the unbelievable task of maintaining the purchasing power of a large number of people who are no longer needed to produce. We may well be on our way to achieving Bertrand Russell's prediction that work will be eliminated from the face of the earth—an idea completely foreign to our forefathers!

These scientific advances have been obtained through the efforts of research workers, carefully and painstakingly investigating the world in which we live. It is estimated that modern American industrial and governmental agencies now spend over $13 billion a year in scientific research and development—a fourfold increase in the last decade. In general, the benefits have been proportional to the outlay, a proposition which, though to be expected, is nonetheless depressing since the outlay has been almost exclusively for research in the physical realm—that is, in such areas as warfare, material inventions, travel, and communication. Research in the social sciences has been neglected despite the obvious lag about which, to paraphrase Mark Twain, everybody talks but nobody does anything. Our achievements in the area of "social technology" have certainly not matched those in the material fields.

We are at the threshold of an age in which science will really come into its own: an age in which we will be able to cross the Atlantic in an hour or less; in which radar-guided cars will travel super-highways at tremendous speeds with a high degree of safety; in which two-way "Dick Tracy" wrist radios will become reality; in which homes will be air-conditioned by thermo-electric panels; and in which weather will be

predicted from information from weather satellites revolving in orbit. In the human aspects, we are entering an era of the transplanted cornea, the synthetic arteries, and, perhaps, the mechanical heart. At the cultural level, microfilms and microcards will probably replace many of the books occupying stack space in our libraries. Closed-circuit TV will place the full facilities of regional libraries in the living room of modern homes; while data retrieval systems will permit scholars and scientists to locate data on any system of classification by the simple press of a button. We can expect to see during our lifetime greater and better things, undreamed of today, as standard equipment.

Lag in Educational Gains

Unfortunately, nothing so spectacular appears likely in the social sciences, but there again the answer will come from research. Whether or not in the year 2000, children come to school in atomic-powered cars only to be taught by methods of pre-World War II vintage will depend on our realization of the necessity of research and on our willingness to bring to the problems of the classroom the same competence that has characterized American industrial research and practice.

This is not to say that no progress has been made in the social sciences. Certainly, science has produced extensive changes in our social and psychological as well as in our economic lives. Science has also given us valuable insights into some of our social problems. Significant gains have been made in education: witness the obvious differences between the classrooms of today and those of the beginning of the century. But we in education have not joined in the spectacular progress which has characterized the physical sciences in recent years. In fact, it seems that many of science's contributions to the social fields have resulted in greater problems and greater difficulties, such as war, delinquency, and divorce. Certainly we cannot blame these difficulties on scientific developments, but these are problems for which science has yet to provide answers. Thus far we have developed a technology of material things; now we need to devote our efforts to the development of an equally adequate technology of human living.

ROLE OF RESEARCH

The Nature of Research

Although most of us recognize that the progress which has been made in our society has been largely the result of *research,* we do not have an exact definition of the term. Most of us have a vague idea of what is involved, but our concept of research generally is too much oriented toward experimentation as conducted in the physical sciences.

Actually research is simply the process of arriving at dependable solutions to problems through the planned and systematic collection, analysis, and interpretation of data. Research is a most important tool for advancing knowledge, for promoting progress, and for enabling man to relate more effectively to his environment, to accomplish his purposes, and to resolve his conflicts. Although it is not the only way, it is certainly one of the more effective ways of solving scientific problems. For our purposes, we can define educational research as the systematic and scholarly application of the scientific method, interpreted in its broadest sense, to the solution of educational problems; conversely, any systematic study designed to promote the development of education as a science can be considered educational research.

Our culture puts such a premium on science that the terms *science* and *scientific* are frequently misused. *Research* is also frequently used in contexts where little research, in the true sense of the word, is actually done. A person no longer looks up a word in the dictionary or a historical fact in the encyclopedia, he "researches" it. Many agencies claiming to do research are engaged in nothing more than fact-finding. In the social sciences, the application of the term *research* should be restricted to activities designed to promote the development of a science of behavior. The term *educational research* should likewise be restricted to systematic studies designed to provide educators with more effective means of attaining worthwhile educational goals. It does not include the routine activities of applying what is already known or of teaching in the usual sense of the word, but is reserved for activities designed to discover facts and relationships that will make the educational process more effective.

Rationale Underlying Research

Research is oriented toward the discovery of the relationships that exist among the phenomena of the world in which we live. The fundamental assumption is that invariant relationships exist between certain antecedents and certain consequents so that, under a specific set of conditions, a certain consequent can be expected to follow the introduction of a given antecedent. Thus, under usual conditions, throwing an object out of the window of a tall building will result in its falling with accelerating speed, and this will be as true tomorrow as it is today, and as true in Brooklyn as it is in Hong Kong. That this invariance in time and space prevails is logical; denying its existence would mean subscribing to a view that phenomena are haphazard, capricious, chaotic, and, consequently, that research and science are impossible.

From the beginning of time, man has noted certain regularities among the phenomena and events of his experiences and has attempted to devise laws and principles which express these regularities. These laws or principles are, of course, not without exceptions; any law is valid only under the conditions under which it was derived. Even though objects tend to fall, they have been known to rise when other forces were active, but this does not deny the general principle of gravity. Research is devoted to finding the conditions under which a certain phenomenon occurs, and the conditions under which it does not occur in what might appear to be similar circumstances. Science is based on the premise that if a given situation could be duplicated in the entirety of its relevant aspects, the phenomenon would also be duplicated without fail. To the extent to which the situation is duplicated only in part, however, the phenomenon may or may not be duplicated. This complication can be frustrating to the beginner who expects to have pat answers to complicated problems, who is interested in *the* solution, without all of the *if's, when's,* and *but's.*

Research and the Social Sciences

Man has come to place more and more reliance on research for the answers to his problems, and, judging from his success in the physical sciences, it appears this trust has been well

placed. To be questioned, however, is the extent to which a similar reliance can be placed on research in the social sciences where, because of the complexity of the variables under consideration, exceptions are more frequent and often are more drastic and damaging. In fact, the many exceptions to the generalizations of the social sciences have led some people to question whether a human science is possible. The point is frequently made that human behavior is not subject to scientific determination and control; that because man is possessed of a free will and is capable of choice, he operates independently of the forces surrounding him.

This view is difficult to accept for the very reason that it denies the concept of order in human behavior. It postulates that human behavior is inexplicable, unpredictable, and uncontrollable, a proposition which even the layman would reject, for he is constantly making interpretations and predictions of the behavior of his fellowmen. It is true that human behavior is relatively complex, and that, from our present level of understanding, it frequently appears disorganized and self-contradictory, but that is simply because we do not understand it. Our present inability to understand does not mean, however, that human behavior is not capable of determination any more than prehistoric man's inability to understand his physical world constituted proof that physical phenomena could never be brought under scientific law.

We might go further and suggest that it is not so much that human behavior is complex as it is that our understanding of human behavior is extremely limited. Our position today with respect to the science of human behavior is not unlike that of the chemist or the geologist at the turn of the century. Rocks and fossils when viewed individually are as different as human beings; yet they have been brought under law. It is likely that the differences among human beings and their behavior patterns lie largely at the peripheral level, and that, when viewed in perspective, significant underlying regularities, uniformities, and generalities in human behavior will begin to appear. In fact, a great deal of relatively definite knowledge of human behavior has already been accumulated in the science of psychology, and progress has been made in inte-

grating these uniformities into a systematic framework and in predicting and controlling the outcomes.

Much of our present knowledge regarding human behavior exists at the level of empirical explanation, and we are now beginning to orient our knowledge of its nature to the task of its prediction and control. This, of course, does not deny that, because of the complex nature of human beings and the way their differences are revealed in their behavior, the prediction of this behavior will be subject to numerous exceptions, and, further, that the laws that are derived will probably never be as simple as those which govern the physical elements. Nevertheless, human behavior is as legitimate a subject for scientific investigation and determination as are the phenomena of the physical world. It might even be suggested that research in human behavior must follow the same basic principles of science, though it may not necessarily subscribe to the same methods as do the other areas of scientific investigation.

Also to be questioned is the extent to which educational research represents a unified field of endeavor of sufficient homogeneity to warrant a common term and common training on the part of those who engage in it. Or, is education so broad and complex that it requires a variety of research techniques which are most appropriately taught in separate courses? The complementary question is whether a single course in research in the social sciences (including education, human relations, psychology, and sociology) might make better sense.

While there are many answers to these questions, it would seem that, though the techniques involved in the various social sciences are based on the same principles, the differences in application of these principles often warrant different emphases and hence separate courses of training. The differences in the various areas of education, on the other hand, appear relatively less crucial, and if education is to make a concerted attack on its problems, a common course should be recommended. This common course would have to be supplemented by special emphasis on the research on the specific problems of the various areas of specialization within education.

Implied in the preceding paragraph is the fact that, in view

of the infinitely more complex phenomena with which social research, in general, and educational research, in particular, have to deal, we must be prepared to adapt our research techniques to the problems with which we are faced. Indeed, it seems likely that much of the difficulty encountered in organizing the social sciences has stemmed from our attempt to mimic the research procedures of the physical sciences. To the extent that the techniques we have borrowed are relatively inappropriate to the problems for which they have been used, the answers have been correspondingly inadequate, and our progress toward the development of education as a science has been correspondingly delayed.

It must be realized that educational research as a tool of science is relatively new, dating back perhaps to the turn of the century. Early progress was slow, largely because of the lack of technical know-how and of such tools as statistical procedures and tests of intelligence and achievement. In fact, it may be said that the breakthrough in educational research dates back to Fisher's presentation, in 1935, of his multivariate experimental design, which made possible an adequate attack on the complicated problems characteristic of the social sciences.

Although it may be more difficult to apply the principles of science to social phenomena, these principles apply to both physical and social phenomena with equal force and effectiveness. Furthermore, in contrast to such sciences as geology and astronomy, where, even after a problem has been solved, there is not very much that can be done about controlling its occurrence, the social sciences generally provide an opportunity for control of any situation whose nature is sufficiently understood.

Our task, then, is to discover the uniformities underlying social phenomena so that they can be integrated into a meaningful structure which will eventually permit the prediction and control of consequences. Although this may be a long, complicated process, it seems logical that, in time, we will be able to predict human behavior with acceptable exactness. In fact, with the newer tools available to us from other disciplines, we should be able to make much more rapid progress than did many of the other sciences. This is precisely the task to which we need to address ourselves as we join our colleagues who only

recently were graduated from alchemy, astrology, and bloodletting.

SUMMARY

1. The scientific advances which characterize our modern world in the physical and material areas have been achieved through thorough and painstaking research, which, is the key to scientific progress. Unfortunately, the social sciences, by contrast, have lagged far behind.

2. Man has always noted certain regularities in the phenomena and events of his experience and has attempted to devise laws and generalizations expressing these regularities and uniformities.

3. Research is oriented toward the discovery of relationships that exist among phenomena. It is predicated on the premise that certain invariant relationships exist between certain antecedents and certain consequents, an assumption which must be accepted since the alternative would be to deny the concept of orderliness and lawfulness in the world about us.

4. Scientific laws are valid only under the conditions under which they have been devised. If two situations could be duplicated in their entirety, their consequents would also be invariably duplicated in full. However, to the extent that the conditions postulated in the statement of a law are duplicated only in part in a real situation, the phenomenon may or may not be duplicated.

5. Social phenomena are as subject to scientific investigation and determination as are the phenomena of the physical world. That human behavior should appear complex and relatively unpredictable and uncontrollable is simply a reflection of the inadequacy of our present knowledge.

6. There is a need to adapt research methods to the complexity of the research problems which exist in education.

7. Our progress to date has been slow. If education as a science is to prosper, we need to bring to educational problems the zeal and the competence that has characterized research in the physical sciences.

PROJECTS and QUESTIONS

1. The major project to be associated with a course in educational research is obviously the thesis or dissertation; undoubtedly, nothing can make a course in research more meaningful than the experience of applying its principles in the solution of an actual research problem. Unfortunately, many students will not be writing a thesis. Two alternatives suggest themselves:

   a) A class project which would involve the close co-operation of the members of the class in every phase of its selection and execution, and, if conducted in a nearby school, close co-

ordination—through a steering-committee—with the school personnel.

b) Individual projects selected and planned but not carried out to completion. Each student might be expected to select a topic capable of development into a master's thesis and to write up what might normally be the chapters on the problem, the review of the literature, and the design of the study (including the construction of the instruments and their pretesting on a small pilot group, where special instruments are required).

2. It is essential that the student enrolled in a course in educational research become oriented toward the nature of research through a thorough survey of research studies actually conducted. This is best done by perusing *Dissertation Abstracts* for studies that appear both interesting and worthwhile. The student might be expected to read the microfilm of at least two or three dissertations in order to understand what constitutes acceptable research.

3. a) Make a survey of the basic research materials in education and related fields. Be sure to include new publications as they become available, the earlier books in the field, and the more pertinent general sources and periodicals of interest to educational researchers.

b) Contrast earlier books in research with the more recent on the basis of such factors as the coverage of the field, the general content with respect to emphasis on the different methods of research, and so on.

4. Specifically, what changes have taken place in educational practice since the turn of the century? How many of these changes actually rest on a firm foundation based on research? How many of these changes are relatively impossible to validate empirically?

5. Visualize the classroom of 2000 A.D. What do you anticipate in the area of the physical plant, the school furniture, teaching aids, and especially, teaching procedures. Specifically, how might research expedite and ensure the validity of the changes that take place in educational practice?

6. Trace some of the material changes that have taken place in the last couple of decades in such areas as transportation and communication, ballistics, medicine, and so on. What might be expected in the year 2000 A.D.? What role might education play in this progress and how can it best orient itself to promote and to guide progress in desirable directions?

SELECTED REFERENCES

AUSUBEL, DAVID P. "The Nature of Educational Research," *Educational Theory*, 3 (October 1953): 314–20.

BARBER, BERNARD. *Science and the Social Order.* Glencoe: Free Press, 1952.

Du Vigneaud, Vincent. *A Trail of Research.* Ithaca: Cornell University Press, 1952.
Freeman, Frank N. "The Contributions of Science to Education," *School and Society,* 30 (July 27, 1929) : 107–12.
Kandel, Isaac L. "Educational Research," *School and Society,* 72 (October 7, 1950) : 232.
Kelley, Truman L. " A Defense of Science in Education," *Harvard Teachers Record,* 1 (November 1931) : 12?–30.
———. "A New Development—Mechanized Information," *Educational Research Bulletin,* 40 (October 1961) : 186.
Editor. "A New Development—Mechanized Information," *Educational Research Bulletin,* 40 (October 1961) : 186.
Russell, Bertrand. *Human Knowledge: Its Scope and Limits.* New York: Simon and Schuster, 1948.
———. *The Impact of Science on Society.* New York: Simon and Schuster, 1953.
Scates, Douglas E. "Nature and Need of Educational Research," *Science,* 103 (May 31, 1946) : 657–62.
———. "On Behalf of Educational Research," *Phi Delta Kappan,* 28 (May 1947) : 365–9.
Walsh, J. Hartt. "Education in 2000 A.D.," *Nation's Schools,* 57 (April 1956) : 47–53; 57 (May 1956) : 86–8.

GENERAL SOURCES

No introductory text can ever hope to provide a complete coverage of an area as broad and as complex as educational research. It is, therefore, imperative for the student to consult other sources not only in order to gain adequate insight into the concepts presented, but also to obtain a greater background in areas of special interest. The following is a representative list of the more appropriate general references in the research methodology of education and related fields. Additional sources of a more specific nature can be found in the references at the end of each chapter. The conscientious student will undoubtedly want to make extensive use of these in broadening his perspective of research. He will also want to attain a reasonable grasp of the content of educational research by becoming familiar with such sources as the *Encyclopedia* and the *Review of Educational Research* and the various professional journals.

Ackoff, Russell L. *The Design of Social Research.* Chicago: University of Chicago Press, 1953.
American Council on Education. *Educational Research: Its Nature, Essential Conditions, and Controlling Concepts.* Series 1, Vol. 3, No. 10. Washington, D.C.: The Council, 1939.
American Educational Research Association. *Growing Points in Educational Research.* 1949 Official Report. Washington, D.C.: The Association, 1949.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION AND DEPARTMENT OF CLASSROOM TEACHERS. *The Implications of Research for the Classroom Teacher.* Joint Yearbook. Washington, D.C.: National Education Association, 1939.

ASSOCIATION FOR SUPERVISION AND CURRICULUM DEVELOPMENT. *Research for Curriculum Improvement.* 1957 Yearbook. Washington, D.C.: The Association, 1957.

BANGHART, FRANK W. (ed.). *First Annual Symposium on Educational Research.* Bloomington: Phi Delta Kappa, 1960.

BARNES, JOHN. *Educational Research for Classroom Teachers.* New York: Putman, 1961.

BARR, ARVIL S., *et al. Educational Research and Appraisal.* New York: Lippincott, 1953.

BEST, JOHN W. *Research in Education.* Englewood Cliffs: Prentice-Hall, 1959.

BEVERIDGE, WILLIAM I. B. *The Art of Scientific Investigation.* New York: W. W. Norton, 1951.

COLLIER, RAYMOND O. and ELAM, STANLEY. *Research Design and Analysis.* Second Symposium on Educational Research. Bloomington: Phi Delta Kappa, 1961.

DOBY, JOHN T. (ed.) *An Introduction to Social Research.* Harrisburg: Stackpole, 1954.

FESTINGER, LEON and KATZ, DANIEL. *Research Methods in the Behavioral Sciences.* New York: Dryden, 1953.

GEE, WILSON. *Social Science Research Methods.* New York: Appleton-Century-Crofts, 1950.

GOLDHAMMER, KEITH and ELAM, STANLEY (ed.). *Research Dissemination and Implementation.* Third Symposium on Educational Research. Bloomington: Phi Delta Kappa, 1962.

GOOD, CARTER V. *Introduction to Educational Research.* New York: Appleton-Century-Crofts, 1959.

GOOD, CARTER V. and SCATES, DOUGLAS E. *Methods of Research.* New York: Macmillan, 1954.

GOODE, WILLIAM J. and HATT, PAUL K. *Methods of Research.* New York: McGraw-Hill, 1952.

HILLWAY, TYRUS. *Introduction to Research.* Boston: Houghton-Mifflin, 1956.

LINDQUIST, EVERETT F. *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton-Mifflin, 1953.

LUNDBERG, GEORGE A. *Social Research.* 2nd ed.; New York: Longmans, Green, 1942.

MCCONNELL, THOMAS R., *et al. The Conceptual Structure of Educational Research.* Supplementary Educational Monograph, No. 55. Chicago: University of Chicago Press, 1942.

MCCORMICK, THOMAS C. and FRANCIS, ROY G. *Methods of Research in the Behavioral Sciences.* New York: Harper, 1958.

MADGE, JOHN. *The Tools of Social Sciences.* New York: Longmans, Green, 1953.

MUSSEN, PAUL H. (ed.). *Handbook of Research Methods in Child Development.* New York: Wiley, 1960.

NATIONAL SOCIETY FOR THE STUDY OF EDUCATION. *The Scientific Movement in Education.* 37th Yearbook, Pt. 2. Bloomington: Public School Publishing Co., 1938.

RUMMELL, J. FRANCIS. *Introduction to Research Procedures in Education.* New York: Harper, 1958.

SELLTIZ, CLAIRE, *et al. Research Methods in Social Relations.* New York: Holt, Rinehart & Winston, 1959.

SMITH, HENRY L. and SMITH, JOHNNIE R. *An Introduction to Research in Education.* Bloomington: Educational Publications, 1959.

TRAVERS, ROBERT M. W. *An Introduction to Educational Research.* New York: Macmillan, 1958.

UNDERWOOD, BENTON J. *Psychological Research.* New York: Appleton-Century-Crofts, 1957.

VAN DALEN, DEOBOLD B. *Understanding Educational Research.* New York: McGraw-Hill, 1962.

WILSON, E. BRIGHT. *Introduction to Scientific Research.* New York: McGraw-Hill, 1952.

The greatest invention of the nineteenth century was the invention of the method of invention.

ALFRED NORTH WHITEHEAD

2 The Evolution of Science

From the beginning of time, man has been curious about his environment. Life is full of intriguing phenomena: from the rain, thunder, and lightning which aroused the curiosity—to say nothing of the terror—of primitive man to the many problems of today which have a direct bearing on man's welfare and, perhaps, on his survival. Therein lies the raw material from which science is born.

MAN'S SEARCH FOR TRUTH

The means by which man seeks the answers to his problems can be classified under three broad categories: *experience, reasoning,* and *experimentation.* These three categories are, of course, complementary and overlapping. Experimentation, for instance, is perhaps best conceived as a combination of experience and reasoning. In fact most problems—and certainly most research problems—call for the operation of varying degrees of all three.

Experience

Perhaps the most primitive, and yet the most fundamental, source of the solution to a problem lies in personal experience. Thus, confronted with a sudden flow of water down a

ravine, prehistoric man could have solved his problem and saved his life if he had only remembered that water does not generally stay on hills. On an elementary level he had learned a basic scientific fact: water runs downhill. Experience is generally considered one of the two arms of science; it is a prerequisite—that is, a necessary, if not a sufficient, condition—to intelligent and scientific behavior.

When one has not had personal experience with a phenomenon, the obvious recourse is to consult someone who has. Thus children consult their parents, their teachers, or even their older siblings for answers to problems with which they are not familiar. Throughout the history of science, certain persons became recognized as authorities—that is, there emerged a class of people who were credited with having many of the answers to the problems that perplexed their less enlightened contemporaries. Frequently, these authorities were merely persons of authority or power whose word was law, not because of any great wisdom or communion with truth, but because of prestige derived through strength, birth, wealth, association with magic, or some other form of public acceptance. A few of these authorities attained historical renown: Plato and, especially, Aristotle are still considered authorities on many things that are worth knowing. More recently, the name of John Dewey has been invoked as the last word on what should be done in education.

Closely related to personal experience are custom and tradition, which provide a large percentage of the answers to everyday as well as professional problems. Much of what goes on in the classroom, for example, is justified by "This is the way we have always done it."

Obviously, experience is a fundamental aspect of the foundation on which science must rest. If we did not profit from experience, the path of science would be dull and limited indeed. On the other hand, as a scientific tool in the discovery of truth, experience has very definite limitations which must be fully appreciated. Primary among these is the fact that frequently one has only an inadequate, if not inaccurate, conception of his experiences. In fact, often one is no more clear about what he experiences than were the blind men looking at the elephant.

The pronouncements of authorities also must be accepted cautiously. Although it is true that certain individuals have such wide experience and/or great insight that their advice is sought by many—and profitably so—it must be remembered that no one is infallible, that authorities frequently disagree among themselves, and that even the best and the most competent are not endowed with "the truth, the whole truth, and nothing but the truth." The matter is complicated further by the tendency for authorities to go beyond their areas of competence. Thus actresses advertise soap and a dozen other things while athletes become authorities on razor blades and supersede the medical doctor as judges of the nutritional value of breakfast cereals.

Furthermore, perhaps because of the prestige given by our society to the man of action who can speak with precision and finality, there occasionally arise "prophets" who speak with equally dogmatic authority on a multitude of topics, from politics and national policy to delinquency and education. Obviously, some of these people are men of wide experience and high intellectual ability, but their pronouncements frequently are nothing more than educated guesses—if not mere personal opinion—and must be considered in that light.

It is also likely that a man who is an authority in one era will be surpassed by very ordinary persons in the next stage of cultural development, and that he will become a liability if he is considered an authority past the life of his contribution. As long as people felt that Aristotle had answered all questions, there was no need for seeking more adequate solutions. Since he had "attained the final truth," the acme of scholarship, education, and wisdom up to the Renaissance consisted largely of mastering his answers and techniques of reasoning. In fact, history suggests that it was not prudent to question his conclusions: note, for example, the suppression of such theories as heliocentricity and evolution and the discoveries of Galileo, Copernicus, and other early scientists. In fact, probably the only scholars who were secure in the Middle Ages were the mathematicians, inasmuch as their "discoveries" were unlikely to conflict with any of the established beliefs of the period. Even today, in the U.S.S.R., the writings of Lenin are considered to be *truth,* and again it is not prudent to disagree. Even in democratic countries, courts of law subscribe to the concept of

experts and authority; and in our schools, policy and practice are frequently based on the opinion of the veteran teacher or administrator.

Custom and tradition also must be evaluated carefully. A person cannot investigate everything for himself, and therefore he must rely on the discoveries of others. Furthermore, anything which has stood the test of time can be expected to possess some element of truth, and thus tradition and custom can be considered reasonably dependable. But that custom and tradition are not infallible guides to truth can be seen from the number of classical errors in history—for example, the belief that the world was flat. The phonetic approach to the learning of reading was a similar error which attained widespread acceptance among educators. Nothing seemed more logical; nothing could be more simple: all a child had to do was to learn the sound of his letters and how to blend them into one another and his reading problems were solved. But then educators discovered that this is not the way that people *begin* to read. Other educational "truths" that have had general acceptance at one time or another include the formal theory of mental discipline, the emphasis on drill, and the schedule approach to baby care.

We must not understate the role of experience and authority in the discovery of truth. Both have a crucial role in promoting man's understanding of his world. They must not be considered as sources of ultimate truth, however, but rather simply sources of suggestions and hypotheses to be questioned and subjected to more rigorous test. It might be pointed out in passing that, in general, our culture does not encourage this sort of questioning, particularly on the part of children. Even adults who are "from Missouri" are not always popular. Teachers, for example, expect their students to accept what they are told, and parroting what the teacher or the book has said, preferably verbatim, is generally the royal road to good grades, teacher acceptance, and other rewards. In fact, even at the college level, disagreement with the instructor is rarely encouraged.

Reasoning

A more sophisticated approach to truth is reasoning, which is considered the other of the two arms of science. Prob-

ably the first major contribution to the systematic discovery of knowledge was made by the Greek philosopher Aristotle, who perfected the syllogistic method of *deductive reasoning*.

In its simplest form, the syllogism consists of a major premise based on a self-evident truth or previously established fact or relationship; a minor premise concerning a particular case to which the fact or relationship inescapably applies; and a conclusion. Thus

> All men are mortal;
> John is a man;
> therefore, John is mortal.

Syllogistic reasoning is governed by a number of rules which are generally incorporated in present-day courses in logic. The procedure is based on the concept of internal consistency; it operates on the assumption that, through a series of formal steps of logic, valid conclusions can be deduced from valid premises.

In contrast to deductive reasoning, which consists of going from the general to the particular, *inductive reasoning* proceeds from the analysis of a number of individual cases to a hypothesis, and eventually to a conclusion concerning the general case. As it applies to the actual solution of a research problem, reasoning is both inductive and deductive—that is, it consists of a back-and-forth movement in which experience is analyzed inductively to provide hypotheses whose implications are then studied deductively in order to test their validity.

The role of reasoning as an aspect of science must not be minimized: it is an indispensable tool in its operation and development. On the other hand, its limitations must also be clearly recognized. A conclusion is no more adequate than the premises (major or minor) on which it is based; false premises can lead only to false conclusions. Errors can also stem from violations of the rules of logic. Barring such errors, reasoning can point to new relationships in what is already known, but it cannot derive new truths. It can indicate that two propositions are in conflict, but it cannot tell which, if either, is correct, and reasoning from analogy is generally unacceptable as a method of deriving truth. Thus, the empirical fact that compounds of elements that are individually combustible in an at-

mosphere containing oxygen are themselves combustible, and that compounds of elements that individually are incombustible in oxygen are themselves incombustible would not enable us to reason whether these rules are exception-free—which indeed they are not. In general, the contributions of reasoning to the development of science are three: 1. the suggestion of hypotheses; 2. the logical development of these hypotheses and the planning of a research design for testing their validity; and 3. the clarification and interpretation of scientific findings and their synthesis into a conceptual framework.

Experimentation

The third and most scientifically sophisticated means by which man seeks to discover truth is experimentation, which is generally considered—and perhaps erroneously—to be the scientific method *par excellence*. Chapter 12 will be devoted to a presentation of the various aspects of experimentation, but it might be pointed out here that, in its simplest form, experimentation consists of isolating the effects of the operation of a given factor by assigning that factor to one of two groups which are otherwise equal in all respects. This is, of course, an oversimplified experimental design, which is rarely approached even in the physical sciences, and probably never in the social sciences.

HISTORICAL DEVELOPMENT

Early Contributions

The beginning of science dates back to the beginning of man. Undoubtedly early man discovered a large number of empirical relationships which enabled him to understand his world with varied success. The first recorded attempts at science are those of the Egyptians, who, partially in response to the annual flooding of the Nile, developed the calendar, geometry, and surveying. These achievements were followed by the less concentrated, but nonetheless valuable, contributions of the Babylonian and Hindu civilizations. Then came the Greeks who, with their emphasis upon organization, gave us not only astronomy, medicine, and the Aristotelian system of classification, but also the syllogism as the basis for further deduc-

tive systematization of experience. Despite their overemphasis of the theoretical—at the expense of the empirical—and their neglect of experimentation as the prime source of scientific evidence, the Greeks can be credited with the first systematic approach to the development of science.

The syllogistic approach was the only effective tool of systematic reasoning during the Hellenistic and Roman periods and up to the days of Galileo and the Renaissance. It reached a peak of misuse in the Middle Ages, when, in disregard of the admonitions of Aristotle, it lost contact with basic observations and experience and degenerated into an exercise in mental gymnastics. For example, if the problem concerned the number of teeth in a horse's mouth, the solution was sought through logic rather than through simple observation, an error of which even Aristotle apparently was guilty. To quote Bertrand Russell:

> To modern educated people, it seems matters of fact have to be ascertained by observation, not by consulting ancient authorities. But this is an entirely modern conception, which hardly existed before the 17th century. Aristotle maintained that women have fewer teeth than men; although he was twice married, it never occurred to him to verify the statement by examining his wives' mouths. He also said that children would be healthier if conceived when the wind is in the north . . . , that the bite of a shrewmouse is dangerous to horses, especially if the mouse is pregnant . . . and so on and so on. Nevertheless, classical dons, who have never observed any animals except the cat and the dog, continue to praise Aristotle for his fidelity to observation.[1]

Up to the Renaissance, Aristotle's teachings were considered to be true, relevant, satisfactory, and at once adequate for any and all purposes, and "science" fell to a new low in sterility and futility.

Bacon and the Inductive Method

In the early 1600's, Francis Bacon led a rebellion against what he considered a tendency among philosophers first to agree on a conclusion, and then to marshall the facts in its sup-

[1] Bertrand Russell, *The Impact of Science upon Society*. (New York: Simon and Schuster, 1953), p. 7.

1372

port, in the same way as one does in a debate where presenting a convincing argument in support of a point of view—rather than discovering the truth—is the main concern. He felt strongly that logic could never suffice for the discovery of truth, since "the subtlety of nature is greater many times over than the subtlety of argument,"[2] and that logic began with a preconceived notion and, therefore, biased the results obtained. He posited that if one collected enough data without any preconceived notion about their significance and orientation—thus maintaining complete objectivity—inherent relationships pertaining to the general case would emerge to be seen by the alert observer.

Bacon's contribution to the advancement of science is significant in that it broke the stranglehold of the deductive method whose abuse had brought scientific progress to a standstill. He ushered in a period in which such men as Galileo, Lavoisier, Harvey, and Darwin, rejecting logic and authority as a source of truth, turned to nature for the solution to man's scientific problems. These men did not reject logic, experience, and authority; they simply used them as sources of hypotheses rather than of proof, and insisted on empirical proof for their verification.

Bacon was essentially incorrect in his basic premise that a hypothesis was prejudicial to complete objectivity. This need not be so—if one sets out to investigate—that is, to test a tentative position—and not to prove a point. In fact, many graduate schools insist that the student writing his dissertation or thesis state in exact terms the hypothesis he plans to test. Bacon's method at its best was obviously wasteful; at its worst, it was essentially ineffective. An investigation not guided by a hypothesis is more likely to result in confusion than in enlightenment and generalizations. This often is seen in the student who compiles tremendous amounts of data and asks, "What should I do with them now?" Practical experience suggests that the best thing to do is to throw them away, since data collected in this way will rarely solve anything. In general, data should be collected for research purposes only after the problem has been clarified sufficiently to suggest a hypothesis worth exploring. The point is clearly made by Larrabee:

[2] Francis Bacon, *Novum Organum* (trans., New York: Willey, 1944).

6.10.93
7460

"Unless he is a mere collector of odds and ends, the seeker of knowledge cannot go through life looking *at* things; he must look *for* some things; that means active inquiry with some directing factor in control."[3]

The limitations of induction in the more advanced stages of science are noted by Einstein, who points out that

> There is no inductive method that could lead to the fundamental concepts of physics. Failure to understand this fact constitutes the basic philosophical error of so many investigators of the 19th century. . . . We now realize with special clarity how much in error are those theorists who believe that theory comes inductively from experience.[4]

The Modern Inductive-Deductive Approach

Bacon's inductive method was superseded by the *inductive-deductive* method—generally attributed to Charles Darwin —which combines Aristotelean deduction with Baconian induction. It consists of a back-and-forth movement in which the investigator first operates inductively from observations to hypotheses, and then deductively from these hypotheses to their implications in order to check their validity from the standpoint of the compatibility of the implications with accepted knowledge. After revision where necessary, these hypotheses are submitted to further test through the collection of data specifically designed to test their validity at the empirical level.

This approach is the essence of the modern scientific method and marks the last stage of man's progress toward the derivation of empirical science—a path that took him through folklore and mysticism, dogma and tradition, unsystematic observation, and finally to systematic observation. Although, in practice, the process involves a back-and-forth motion from induction to deduction, in its simplest form, it consists of working inductively from experience to hypotheses, which are elaborated deductively for implications on the basis of which they can be tested. Thus the modern scientist does not rely exclusively on induction, but rather deduces the

[3] Harold A. Larrabee, *Reliable Knowledge* (Boston: Houghton-Mifflin, 1945), p. 167.

[4] Albert Einstein, "Physics and Reality," *Journal of Franklin Institute*, 222 (1936) : 349–89.

implications of his hypotheses in relation to what is already known. He uses facts and theories as interdependent tools for greater scientific insight into his problem. This dual approach is necessary because, though a scientist is interested in a class rather than in the individuals who make up this class, he can never observe the whole class and must, therefore, generalize from a few instances concerning properties belonging to the whole class.

THE DEVELOPMENT OF SCIENCE

Animism

Man's task is essentially that of understanding the phenomena which he encounters in order to achieve the means of dealing more effectively with the problems which they pose. Primitive man, hearing thunder and seeing flashes of lightning accompanied by violent rain and, perhaps, floods, must have spent many anxious moments wondering when it would all cease and what it was all about.

Anthropology and history indicate that man first explained such phenomena as the work of gods, spirits, demons, and other supernatural agents. Ancient mythology is full of gods and goddesses who obviously played a significant part in the lives of these primitive people. The Indians attributed sickness, famines, and other misfortunes to the displeasure of the spirits. Even today the ceremonials of primitive tribes are designed to appease angry spirits or to secure their help. This supernatural or animistic stage is not altogether past even among civilized groups. It is not uncommon for "modern" people to believe in ghosts, gremlins, Lady Luck, and various other beings invented to explain phenomena of unknown origin. Irish folklore is particularly filled with such myths, and even in our country such superstitions as black cats, ladders, Friday-the-13th, and "hexing" with voodoo dolls is still probably more prevalent than we as a civilized nation would like to admit.

Empirical Science

In time, man came to realize that natural phenomena can be explained on the basis of natural *causes*—a most important

step marking the beginning of science as a systematic approach to the solution of problems. This development has been a slow process. Crude, unsystematic conjectures gradually gave way to more systematic and critical observation; then to systematic and precise testing of isolated hypotheses under controlled conditions; and finally, in some sciences at least, to the development of theories incorporating the findings of isolated "experiments" into an integrated structure, and to the formulation of systematic and precise tests of integrated hypotheses derived through such theories. This process can be divided into two overlapping stages of development: 1. *the empirical level,* in which science consists of discovering empirical relations among phenomena of the variety of "X leads to Y," without understanding *why* this is so; and 2. *the explained* (theoretical) *level,* which involves the development of a theoretical structure that not only explains isolated empirical relationships, but also integrates them in a meaningful pattern. This theoretical level represents the most advanced stage of science, a stage which is not completely attained in any of the academic disciplines, and which is, of course, barely approached in the social sciences.

1. *Experience.* Obviously the starting point of science at its most elementary level is experience, whether the phenomenon be a thunderstorm, a snowstorm, a vessel broken as a result of the expansion of water as it freezes, an eclipse, or the more commonplace regularity of day and night. Science starts with an observation to which are added other observations both of a similar and a dissimilar nature, until similarities and differences are identified. Eventually a system of basic principles is derived that will explain both the occurrence and the non-occurrence of a given set of experiences. The goal of science is the acquisition and the systematization of knowledge concerning the phenomena experienced.

In its early stages, science must concern itself with augmenting and criticizing experience. The accumulation of individual experiences, no matter how clarified, however, is not sufficient, for as long as these experiences remain isolated, they tend to have no meaning from the standpoint of science. The number and diversity of these isolated experiences must be reduced to their underlying unitary basis of organization

through the process of classifying and systematizing them into a small number of basic principles of ever-greater generalizability and application.

2. *Classification.* The most basic procedure for reducing isolated data to a functional basis is classification, a procedure which is fundamental to all research—and to all mental life—for it constitutes a simple and parsimonious way of comprehending large masses of data. Knowing the class to which a given phenomenon belongs provides us with a basis for its understanding. The classification of a forthcoming storm as a hurricane, for example, gives us the basis for anticipating much of its probable behavior, for the identification of an object or phenomenon as a member of a class immediately implies certain attributes already associated with the class. Thus, the words *fish*, *bird*, and *diamond* carry certain meanings with them. And, the more precise the classification, the clearer the meaning and the more specific the properties associated with the classification. The properties associated with the class *robin* are more precise than those associated with the general class *animals*, for example.

To be meaningful, classification must be based on one's purpose. Thus, whether an orange should be classified with a banana or with a baseball depends on whether one is interested in eating it or in rolling it along the floor. Complications arise from the fact that most objects and phenomena have a large number of properties and characteristics, and thus can be classified in many ways. For example, though telephone directories are organized by surname and by business affiliation, they could also be organized by street addresses. The problem is to distinguish between a crucial and a superficial basis of classification. Children, for instance, might classify cars by their color rather than by their more fundamental properties. In fact, it is characteristic of the early phases of classification for it to be based on superficial properties; more crucial bases emerge only gradually as greater insight into the phenomena in question is derived. Thus from the standpoint of the effect a teacher has on his students, the classification of teachers into autocratic and democratic is probably more meaningful than their classification according to the more tangible factors of age, sex, or degree status. Indeed, the fact that individual differences exist among the

members of a given class with respect to almost any trait—at least in the social sciences—is evidence of a lack of complete or perfect classification.

Classification systems can range from the most simple to the most complex and elaborate—perhaps involving multiple bases of classification or even a classification within a classification.[5] Probably the most advanced classification system in existence is the periodic table in chemistry which dates back to the Greeks' early classification of elements into earth, air, water, and fire. Another major classification system is that of plants and animals, a system which, despite its high level of refinement, becomes somewhat difficult to apply in the case of marine animals. More closely related to the present text is The Library of Congress (or the Dewey-Decimal) system of classifying library material.

The actual allocation of objects or phenomena to different classes—whatever the merit of the classification system may be —is relatively easy if their properties can be appraised and if the basis of classification is known in advance, or if agreement can be reached about the classes where such classification is arbitrary. Since a frog is a frog and a lizard is a lizard because they have the properties of a frog and a lizard, respectively, correct classification poses no problem. Similarly, people can be classified as tall, average, and short, if we agree on the limits of these classifications.

3. *Quantification.* While the first step in the development of science is the accumulation and clarification of experiences, it soon becomes necessary to quantify these observations for, though qualitative observations may be satisfactory in the early stages of a science, only quantification can provide the precision necessary for classification in a more mature science. In fact, the more advanced the science, the greater is the need for it to go beyond enumeration toward ever-greater precision in measurement in order to permit the more adequate analysis of phenomena through mathematical manipulation. On the other hand, one must not lose sight of the fact that, though

[5] The modern emphasis on the concept of *sets* in mathematics, which is essentially a system of mathematical logic based on classification—for example, the set of odd numbers or of prime numbers—is of interest in this connection.

quantification permits an infinite set of fine distinctions, mathematical refinement does not endow data with a precision and a significance they did not possess in the first place.

4. *Discovery of Relationships.* Through the classification of phenomena along different continua, it is frequently possible to note certain functional relationships among their component aspects. Classifying children simultaneously by sex and physical strength, for instance, is likely to emphasize the fact that boys tend to be stronger than girls. Functional relationships among phenomena can also be observed through temporal sequences. For example, hot days tend to be followed by electrical storms and showers. At a more advanced level, empirical science attempts to express natural laws in the form of a numerical equation relating the quantitative aspects of certain variables to those of other variables—for example, $c = 2\pi r$ or $V = IR$.

Many of the relationships so discovered represent nothing more than functional co-appearances among phenomena. Such relationships are often crude and indirect. Thus, the relationship between the physical size of an unselected group of children and their reading ability is a spurious relationship: the more correct version is that physical size is related to chronological age which is, in turn, related to mental age and to reading competence. Many similar examples can be given—for instance, the person who got intoxicated on water and gin, on water and bourbon, and water and rye, and who decided that water, being the common element, was the cause of his difficulty. The fact that the mortality rate among inmates of jails and penitentiaries is lower than that of the nation as a whole can be explained on the basis of the age bracket of the people incarcerated. By the same token, an increase in delinquency may simply reflect stricter law enforcement. Another classic example of a similar error concerns the Hawthorne experiments[6] in which what was compared was not the relative intensity of illumination in the factory, but the relative effects on production of attention and neglect.

[6] See Chapter 15 for a discussion of the Hawthorne study and others used to illustrate some of the points in this and the other chapters of the text. The student not familiar with these studies would do well to give Chapter 15 a cursory reading.

5. *Approximations to Truth.* Scientists are generally interested in more fundamental relationships than those representing mere concomitance. Events are frequently so complex that any relationship that may exist between them is blurred. It is necessary, therefore, to analyze them into their basic constituents with the object of discovering a more precise relationship. Accordingly, a major aspect of science consists of the analysis of phenomena to determine more clearly the relevance of their many aspects.

Involved here are two very fundamental steps in the development of science: the process of successive approximations to the truth, and the parallel process of redefinition of the problem in the light of the success or failure of such approximations. This has been particularly evident in recent years in connection with the polio vaccine, where we have had, in succession, the Salk, the Coxe, and the Sabin vaccines, each looking like *the* solution only to be found wanting. Numerous examples of the kind can be found in any scientific field—in agriculture, for example, better varieties of wheat and other grains are discovered every year. Whether in the end, as far as it pertains to natural phenomena, the ultimate truth can be attained is a matter of debate; the issue is relatively academic and probably serves no useful purpose. In many cases, we have come to an approximation to the truth which is sufficient to our purpose—for example, the immunization against smallpox.

The concept of science as but a series of approximations to the truth, which is but rarely, if ever, attained, is not particularly satisfying to those who conceive of science as something absolute and who fail to appreciate that all that science can do is provide us with greater understandings. Of interest in this connection is the relatively prevalent tendency, say, in the field of medicine, to use a "shotgun" approach. The patient is given a general drug, such as penicillin, which may bring about a recovery but which, since it does not help identify the curative agent, does not provide the basis for the future treatment of similar cases, except for the repetition of the general approach. To be of maximum scientific value, the approach should be to try one drug at a time or, if it were possible to obtain a sufficient number of cases, to try a variety of drugs in

combinations of one, two, three, and so on in a more elaborate experimental design.

Theoretical Science

The ultimate level of science is *theoretical* or *explained* science, in which the relationships and phenomena discovered in empirical science are explained on the basis of underlying causation as a step toward predicting and determining the methods of controlling their operation so that desirable outcomes can be promoted. This advanced stage of science is more likely to be attained in the physical sciences than it is in the social sciences. For a long time, for instance, chemists realized that certain substances burned, giving up heat and smoke, and leaving ashes. This was worth knowing in itself, but it did not explain what was occurring. They proposed various theories to explain the occurrence, among which was the postulation of a substance they called *phlogiston* present in the atmosphere and apparently responsible for the burning. This theory was later rejected in favor of the modern oxidation theory which relates the burning of wood to the rotting of wood, the rusting of iron, and other chemical reactions.

The superiority of the theoretical over the empirical level of science is best understood through an appreciation of the limitations of the latter. Empirical science is awkward and unwieldy since it deals with phenomena in relative isolation, and thus entails the relatively impossible task of understanding and remembering each phenomenon. Empirical science is particularly limited from the standpoint of predictability and control, which are the ultimate purposes and goals of science. Take the story of little Bobo who, as a result of having accidentally set fire to his hut, was apparently the first human being to taste roast pork. He had an empirical fact; whatever had happened had provided him with roast pork. However, should he want to taste roast pork again, must he burn the hut again? Must he duplicate all the circumstances that anteceded the roast pork? Furthermore, his empirical findings might lead him astray in believing that he could also improve the taste of rice by setting fire to the hut. Bobo's empirical knowledge was of restricted usefulness, though, to be sure, he now had a goal that he might pursue. He might have correctly identified the reason

for the roast pork through intuition, or he might have attempted by trial and error to eliminate one factor after another and, eventually been able to simplify the procurement of roast pork.

Theoretical science can short-cut the process of arriving at solutions. When the individual understands the causes of an occurrence he can transfer his knowledge to advantage in the solution of similar problems. Explained science has obvious advantages from the standpoint of stimulating research and providing worthwhile hypotheses. In fact, the ultimate in scientific excellence is found in such sciences as physics, in which theory has advanced sufficiently (on the basis of previous empirical discoveries) that it can now anticipate and lead in the discovery of empirical facts. The atomic bomb, for example, was not devised empirically and then explained; on the contrary, Einstein and his co-workers developed it theoretically, and turned to empirical verification largely for the purpose of eliminating flaws in its execution.

The transition from empirical to theoretical science is, of course, a difficult step. It is relatively easy to find out what occurs, but it is not as easy to explain why. This is particularly true, for instance, in the social sciences where, for example, we still do not have a scientific explanation for the bulk of even the most elementary aspects of what occurs when a child learns. In some of the more advanced physical sciences, considerable progress has been made in this direction, though in none of the sciences is there complete agreement on all aspects. For instance, physics explains the phenomenon of light by two conflicting theories of wave motion and particle movement. In the social sciences, psychology has devised a number of theories to explain psychological phenomena, but none has complete acceptance, nor has anyone provided a complete explanation of all aspects of behavior. We have yet to explain, for instance, the neuro-physiological basis of learning.

As a science, education is almost exclusively at the empirical level. In fact, we have yet to discover many of the empirical relationships that apparently exist among the variables operating in the classroom. Probably our greatest lack, however, is our failure to devise a theoretical framework within which

we can synthesize educational findings made thus far. It might be said that to date the social sciences have suffered from an overemphasis on empiricism and a corresponding neglect of theory. Only recently has there been a realization of the fact that empiricism represents an incomplete stage of scientific development and of the need for greater theoretical orientation.

CAUSATION

Goals of Science

The purpose of science is to establish functional relationships among phenomena with a view to predicting and, if possible, to controlling their occurrence. Of course, even in some of the more developed sciences—astronomy or geology, for example—the impossibility of manipulating the variables[7] restricts us to the prediction of phenomena and, at best, to an adaptation to their occurrence. Nevertheless, it might be highly desirable and useful to predict in advance the likely occurrence of an earthquake so that we can prepare for it, even if we cannot forestall it. It is likewise profitable to anticipate that the dull child will encounter academic difficulties.

Unfortunately, many of the functional relationships that have been established among phenomena are relatively crude and imprecise, incorporating many irrelevant aspects while some of the more crucial aspects go unrecognized, or are only partially understood. Consequently, the resulting predictions frequently have been unnecessarily clumsy and unwieldy, on the one hand, and frequently in error, on the other. This is evident, for instance, in the relationship which has been established between ability and achievement. Furthermore, even if it were possible to recognize all of the factors involved in a phenomena, they could not all be controlled, and prediction still would be inaccurate to some degree. For this reason, the laws of science are always approximate, especially in the social sciences, where the prediction of whether a child will get a cer-

Modern advances in space flight will undoubtedly cause us to reconsider this statement. Already numerous experiments are being conducted of the effects of the tremendous pressure and temperature changes upon bodies projected into space.

tain answer or react in a given way, for example, frequently is determined by some small and unrecognized aspect of the overall situation.

Causation as Probability of Occurrence

Until recently, research was oriented toward the establishment of cause-and-effect relationships among phenomena. Unfortunately the concept of causation has been troublesome to the scientist and the philosopher as well as to the layman. The latter is likely to consider any antecedent or concomitant of a situation as its causative agent. As Bertrand Russell pointed out, the people of every country attributed the depression of the thirties to the sins of their own governments, and, as a result, there was a movement toward the right where the government in power had been leftist and toward the left where the government had leaned to the right.[8] The modern scientist, on the other hand, is more aware of the difficulties the concept presents, and the term *causation*, in its strict sense, is gradually disappearing from the vocabulary of science. The present view was anticipated by Russell, who in 1929 wrote:

> All philosophers of every school imagine that causation is one of the fundamental axioms of science, yet, oddly enough, in advanced sciences, such as gravitational astronomy, the word "cause" never occurs . . . . the reason why physics has ceased to look for causes is that, in fact, there is no such thing. The Law of Causality, I believe, like much that passes among philosophers, is a relic of a bygone age, surviving . . . only because it is erroneously supposed to do no harm.[9]

The refinement of functional relationships into their minimal essentials and, of course, their theoretical explanations are much more difficult than the mere establishment of these relationships. The difficulty stems from the fact that phenomena generally occur as a result of multiple causation, each cause contributing to their occurrence as a vector force both singly and in interaction with others—for example, intelligence contributes only indirectly to teacher effectiveness.

[8] Bertrand Russell, "On the Importance of Logical Form" in Otto Neurath, *et al., International Encyclopedia of Unified Science*, Volume 1, No. 1. (Chicago: University of Chicago Press, 1938), pp. 39–41.

[9] Bertrand Russell, *Mysticism and Logic* (New York: Norton, 1929), p. 180.

Conversely, the occurrence of a given phenomenon as anticipated is a function of the simultaneous operation of all the contributing forces exactly as postulated. However, inasmuch as the fulfillment of the latter condition in a given situation is relatively a matter of chance, the emphasis in research, particularly in the social sciences, has shifted toward the discovery of functional relationships which can be expressed as probability of occurrence. This is particularly true in those sciences in which it is not possible to manipulate variables so that relevant factors can be isolated from irrelevant factors.

Even in sciences where variables can be manipulated, it is almost impossible to control all the factors of a situation to the point of identifying the causal agent or agents, and to preclude with certainty the operation of extraneous factors and thus to guarantee the occurrence of the cause-effect sequence exactly as postulated. Science is now reconciled to the idea that all that can be expected in the situational realities under which science must operate is prediction—and eventual control—at a high level of probability. It must, of course, be realized that the establishment of causation is not essential. Thus we can predict that learning will take place even though we cannot identify its "causes," and we can predict the movement of the planets though we cannot control such movements. This is not to minimize the desirability of establishing rigid cause-and-effect relationships, if this were possible, since such relationships are much more conducive to the development of control and complete explanation than is simple concomitance. It is simply a recognition of the fact that science rarely has complete insight into the nature of such relationships, and that it is invariably incapable of precluding all influences that can vitiate the prediction. It is a recognition that, at best, science can only approximate what might be the ideal statement of the relationship between an effect and its cause or causes.

Mill's Canons

Probably the most systematic of the early analyses of causation was presented by John Stuart Mill,[10] who, in the mid-nineteenth century, postulated five basic canons governing the identification of causes and effects. Despite their obvious

[10] John Stuart Mill, *A System of Logic* (New York: Harper, 1873), Ch. 8.

limitations, which have led to their relative rejection by modern scientists, Mill's canons constitute one of the major advances of scientific thought. In fact, their contribution to the clarification of the all-important concept of causation probably puts them on a par with that of Aristotle's syllogism in the logic of science.

The best known and, in a sense, the most basic are his first two canons.

The Method of Agreement

> If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause (or effect) of the given phenomenon.

Thus, if the people who fell sick after having attended a certain banquet have in common the fact that they ate the same food, it may be suspected that the food is the cause of their illness.

The Method of Difference

> If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.

These canons have been subjected to numerous and thorough critical reviews. Cohen and Nagel[11] point out that they cannot be used either as instruments of discovery or of proof. For example, using the method of agreement in the case of baldness, it would be impossible to find two men different in every respect except that both are bald. If we modify the statement to include only relevant respects, we would have to decide what is relevant, and thus would have to start with a hypothesis about the likely causes of baldness. Using the method of agreement as a method of proof, we would not know whether to expect single or multiple causes and, even if we were to locate one single point of difference between two in-

[11] Morris R. Cohen and Ernest Nagel, *An Introduction to Logic and Scientific Method* (New York: Harcourt, Brace, 1934), p. 251ff.

stances, we could not be sure that it would be present in all cases. Then, too, there is the possibility that the crucial factor has been overlooked and is not included in the consideration —for example, in this instance, the existence of an unknown germ which is the cause of baldness.

From the standpoint of scientific rigor, the method of difference suffers from much the same limitations as the method of agreement. As a method of discovery, it is limited by the fact that two circumstances can never be alike in all respects save one. Restricting the statement to relevant circumstances necessitates a hypothesis, which may overlook the agent causing the phenomenon directly, or acting as a catalytic agent in the operation of other variables. Nor can the method of difference be used as a means of proof inasmuch as other factors which are not included may be the crucial agents. The problem is further complicated by the fact that, in practice, causal factors are generally complex rather than single. The factors that make for good teaching, for example, probably operate only in interaction with one another.

In summary, the method of agreement is of relatively little value in research since it is almost impossible to fulfill the required conditions. The method of difference, on the other hand, is somewhat more realistic and is essentially the model of the simple experimental design involving the operation of a single experimental factor. Mill's first two canons display their greatest validity and usefulness when stated conversely: A circumstance which is not common to all instances of a given phenomenon cannot be its cause, and a circumstance cannot be the cause of a phenomenon if, when it is present, the phenomenon fails to occur. The use of this converse can be exemplified in the case of identifying the cause of an allergy by eliminating, one by one, the potential causes to the point that the true cause is finally identified through the process of the elimination of non-causes. It can be utilized, for example, in a trial-and-error approach to the diagnosis of academic difficulties.

The Joint Method of Agreement and Disagreement

> If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances

> in which it does not occur have nothing in common save the absence of that circumstance; the circumstance in which alone the two sets of instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.

The third canon is simply a combination of the first two. It might be illustrated by the fact that one event preceding another does not prove that the first is the cause of the second, unless it can also be shown that the suppression of the first also suppresses the occurrence of the second. In a sense, the third canon comes closer to the concept of cause and effect than either of the other two, if chance fluctuations are eliminated through replication and randomness. On the other hand, as pointed out by Cohen and Nagel,[12] the joint method is not much of an improvement, if any, over its two component canons, for it has essentially the same weaknesses as a source of both discovery and proof.

The Method of Residue

> Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedent.

If ABC is the cause of XYZ, and further, if AB is the cause of XY, then C must be the cause of Z. This canon arrives at the specific cause by the process of elimination and is interesting as a source of hypotheses in the more advanced sciences. For example, chemists measuring a mole of chlorine consistently found its weight to be approximately 35.5 grams, which did not make sense since its atomic weight had been calculated on the periodic table to be 35. The problem was solved by the discovery that common chlorine is really a mixture of Chlorine 35 and its isotope Chlorine 37 in the approximate ratio of 3:1. A parallel situation in education would be the case of the gifted child who, because of the interference of emotional factors, for example, is not living up to academic expectations. Useful as this canon is in a highly developed science, it is somewhat less valuable in the social sciences where the expected values are relatively unknown, and the existence, magnitude, and direction of any deviation from the expected cannot always

[12] Cohen and Nagel, *Ibid.*

be appraised. The canon is useful only to the extent that one can determine that an established relationship is being interfered with.

The Method of Concomitant Variation

> Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner is either the cause or an effect of that phenomenon, or is connected with it through some fact of causation.[13]

This canon also has serious weaknesses. Besides requiring valid and reliable measurements, it cannot provide proof since it does not eliminate the possibility of the two factors being caused by a third factor (which is allowed for in the latter part of the statement). Without much effort, it is possible to obtain a correlation of some magnitude between almost any two variables—for example, teachers' salaries and the annual consumption of intoxicating liquors, or the number of teachers in a given community and the number of traffic violations, marriages, and other factors that accompany an increase in population. The value of the concept of concomitant variation probably lies in its negative statement—that is, phenomena that do not vary concomitantly cannot be related causally. Of course, this canon, like the others, can suggest hypotheses for investigation. The failure of the canon of concomitance to isolate the operation of a third factor as the mutual cause of the covariation of the two variables is evident in the Hawthorne studies previously cited.[14] On the other hand, concomitant variation is frequently the only way we can deal with certain problems in which, because of the impossibility of the physical manipulation of the variables, we must rely on statistical analysis.

Evaluation of Mill's Canons

Despite their limitations, Mill's canons represent important landmarks in the history of science, for they mark an important gain in the successive definition and redefinition through which science must pass on its way to greater accuracy and truth. As instruments of present-day science, their use is

[13] Mill, *op. cit.*
[14] See Chapter 15.

restricted largely to the derivation of hypotheses and their elimination through logic and, thus, to the reduction of the number of hypotheses to be tested empirically.

Mill's canons were oriented toward the establishment of causation in the sense of attributing the occurrence or non-occurrence of a phenomenon to the operation of a given factor. Their basic weakness stemmed from overlooking the facts that phenomena generally occur in response to a multiplicity of causes, and that what in practice appears to be the same cause frequently leads to varying outcome with the result that the occurrence of a given phenomenon is a matter of probability rather than certainty. Science is now oriented toward the concept of concomitance and probability, and modern statistical developments in multivariate analysis have made the use of Mill's canons relatively restricted, inadequate, and generally naïve.

PROOF

Nature of Proof

Closely related to cause and effect is the concept of *proof*. The determination of precisely what constitutes proof is sometimes difficult since proof implies the possibility of establishing truth, a point in sharp conflict with the modern view of science as a series of steps toward or approximations to the truth—that is, a parallel series of partial proofs rather than proof itself. In fact, even if truth were attainable, it would be difficult to establish with any degree of confidence that it had been attained in any one instance, and it could never be done with certainty. From an empirical point of view, proof parallels the concept of causation and poses the same problems in its establishment, and again, we find a great deal of looseness in the use of the term. As pointed out by Burton[15] the layman makes comments without proof and does not expect proof of the comments which he hears; he neither realizes the necessity for proof nor understands the nature of conclusive proof.

It is first necessary to realize that there are different kinds of proof. For example, in the legal sense, the accused may be considered "proved" guilty when his guilt is shown beyond rea-

[15] William H. Burton, *et al.*, *Education for Effective Thinking* (New York: Appleton-Century-Crofts, 1958), p. 142.

sonable doubt. Corroboration of adverse testimony, preponderance of evidence, failure to establish an adequate alibi, and so on, are generally adequate for conviction. An important consideration is the general reputation of the key witness. However, legal "proof" generally requires the establishment of a motive as the center of a conceptual scheme within which the offense can be viewed in perspective. This tends to be necessary even when the accused postulates his guilt.

Deductive Proof

Of greater interest here is mathematical proof. In geometry, for instance, each theorem ends with the usual Q.E.D. Analysis of such proofs shows them to be based upon internal consistency. Each theorem follows logically from the premise established in its statement and from the proof of the preceding theorems, all the way back to Theorem One, which derives its proof from self-evident postulates and axioms.[16] Barring errors of deductive reasoning, each theorem is as sound as the theorems on which it is based. Proof in mathematics, since it is based on specific assumptions, is much more rigorous than proof in the empirical sense, but its applicability is restricted to the situation specified in its premises. This type of deductive proof is also found in logic, in both cases, proof is absolute: if we accept the premises, barring errors in the process, the conclusions that follow are indisputable.

Empirical Proof

The proof with which research is most directly concerned is empirical proof. What constitutes proof that teaching by Method A leads to greater pupil gain than does teaching by Method B? Or that treating wood with Chemical X increases its tensile strength? Here the proof is relatively more complicated. In the simplest case of the latter problem, one might take two identical boards, impregnate one of the boards with Chemical X and leave the other untreated, and then test both pieces for strength. The essence of proof in empirical science consists of empirical observations confirming a given hypothe-

[16] It is evident that Euclidian geometry, for example, is not so much an expression of mathematical reality as it is precisely what one gets when one starts out with Euclidian assumptions. The development of non-Euclidian geometry testifies to the possibility of ending with a different product if one starts with different premises.

sis. Thus, if the investigator hypothesizes that Chemical X increases the strength of wood, and if, when he subjects his hypothesis to a test, his observations are in line with that hypothesis, proof seems to have been established. Such proof is supplied by nature; the investigator simply notes the answer that is provided.

In practice, establishing empirical proof is invariably complicated since it requires the establishment of control in an attempt to preclude the operation of extraneous factors. Obviously all biasing factors must be controlled since these would vitiate any proof derived. However, even after reasonable control of biasing factors has been effected, the investigator must still consider the operation of chance factors. These are generally controlled through replication and randomization so that whatever effects such uncontrolled factors have on the operation of the factor under study will tend to neutralize themselves—at least at a level that can be estimated and allowed for. This means, of course, that empirical proof, even at best, is always a matter of probability, never certainty.

Proof in Modern Science

The present orientation of science toward probability of occurrence is undoubtedly the most crucial distinction between modern science and that of years ago. Until recently, knowledge was conceived as precise and unalterable, and scientific effort was directed toward the derivation of immutable laws of a cause-and-effect nature. Modern scientific thought, in contrast, subscribes to the concept of general regularities on an overall basis. In other words, there has been a shift from a subscription to such concepts as *truth, cause, proof,* and *mathematical impossibility* to those of *statistical probability* and *improbability.* Implied here is the very real possibility—in fact, the general expectancy—that values comply with a given law only in the general sense, that science is a matter of general regularities rather than exception-free relationships expressed in precise mathematical formulas.[17]

[17] To be noted in this connection is the vital contribution of such statisticians as Gauss and LaPlace in placing the concept of error under law (the *law of error*), and thus permitting the interpretation of data on the basis of their relative agreement with the general law as postulated.

Acknowledging the concept of probability as the basis of scientific proof is an admission that the last word of science has yet to be said. It attests to the incompleteness of our knowledge and the inadequacy of our control of extraneous factors. The fact that we cannot predict certain phenomena with precision implies that we are not controlling all of the factors involved in their occurrence, just as the occurrence of an event from several alternative causes indicates that we have not defined the exact nature of the cause or of the effect. It must not be inferred that the conception of scientific knowledge as merely probable is a denigration of science. Rather it is a recognition that this is the only kind of knowledge possible, even in the most advanced sciences, where reality insists that the actual occurrence of a phenomenon is not in precise agreement with the laws describing its operation. Boyle's Law, for example, expresses an ideal situation which is only approximated in practice. Two missiles fired from the same gun under "identical" conditions rarely, if ever, follow the same trajectory; landing on a target is strictly a matter of probability. For the same reason, actuarial laws, despite their overall accuracy, are almost completely inapplicable in the individual case, just as prediction of academic success is, and must remain, a group concept. On the other hand, there is no cause for discouragement: barring gross errors, parts are manufactured, bridges are built, and satellites orbited. We merely acknowledge that they do so not according to exact specifications, but within the tolerance limits postulated by the law of chance.

It is interesting to note that mathematics, as a pure discipline, is independent of empirical proof or disproof. That two and two make four is not proved or disproved by showing that two oranges plus two oranges make four oranges or that two gallons of sand plus two gallons of water do not make four gallons of anything. Nor do mathematicians care: when practice confirms theory, it is to be expected; when it does not agree, it simply reflects failure in the fulfillment of the assumptions on which the mathematical expressions rest. Nevertheless, it is worth noting that the principles of mathematics are generally supported by the empirical evidence. For example, the weight that can be supported on a clothesline is a function of the angle of declination and the tensile strength of the line

as represented by the equation $w = t \sin \theta$. This, of course, does not "prove" the validity of trigonometry as a discipline, since mathematical proof is not empirical; it simply "proves" that the assumptions upon which trigonometry is based are apparently sound and that the results are therefore useful.

SUMMARY

1. Man's attempts to arrive at truth concerning his environment have been based on three complementary approaches—*experience, reasoning,* and *experimentation.*

2. Experience—our own or that of others—obviously is a prerequisite to the development of science. However, the number of false beliefs attest to its limitations as a source of scientific truth. Experience makes its greatest contribution in the derivation and the verification of hypotheses.

3. Reasoning is an indispensable tool in the derivation of scientific truth, but, it cannot generate or even identify truth. Its contribution to the development of science lies in suggesting hypotheses, in evaluating the compatibility of hypotheses with accepted knowledge, in devising a research design capable of testing these hypotheses, and in interpreting the results of such a test.

4. Historically, syllogistic reasoning, perfected by Aristotle, represents the first systematic attempt at the discovery of truth. It degenerated during the Middle Ages into an exercise in mental gymnastics, divorced from basic experience. It was superseded in the early 1600's by the inductive approach advocated by Francis Bacon, who objected to what he considered to be the prejudicial influence of hypotheses in orienting the scientist to a prejudged conclusion. Bacon's approach—which was undoubtedly wasteful, if not unproductive—was, in turn, superseded by the modern inductive-deductive method, generally credited to Charles Darwin.

5. Experimentation is undoubtedly the most rigorous approach to scientific truth. It is designed to test the validity of hypotheses under rigorous conditions of control.

6. In its development, science went through three relatively clear, although overlapping, stages. Primitive man explained phenomena on the basis of gods, spirits, and other supernatural agents. Later, man came to realize that natural phenomena had a natural cause, and eventually he undertook the task of deriving empirical generalizations relating the occurrence of phenomena to their cause. The third stage consists of developing a logical framework to explain the empirical relationships noted and to permit the deduction of hypotheses concerning the other aspects of the phenomenon in question. This represents the highest level—and the goal—of scientific endeavor.

7. Empirical science operates through such steps as: (a) the

accumulation and clarification of experience; (b) classification; (c) quantification; (d) discovery of relationships; and (e) successive approximations to (and successive re-definitions of) the truth.

8. Early science—as exemplified by Mill's Canons—was oriented toward the discovery of cause-and-effect relationships of the one-to-one variety between a certain antecedent and a certain consequent. Modern science, in contrast, recognizes that a multiplicity of interacting "causes" are involved in the occurrence of a phenomenon, and that its actual occurrence as anticipated is a function of the simultaneous operation of all contributing factors exactly as postulated, and, therefore, always a matter of probability, never certainty.

9. What constitutes proof needs clarification. Deductive proof—like that in mathematics and logic—is simply a matter of internal consistency. Legal proof is generally a matter of plausibility and general credibility. Empirical proof, on the other hand, is invariably a troublesome concept, since it implies the establishment of cause-and-effect relationships (or truth)—which, as we have seen, is relatively impossible. Modern science is oriented toward successive approximations to the truth and partial (tentative) proofs. In contrast to the earlier view, modern science views empirical proof—like causation—as a matter of general expectancy or probability of occurrence at a given level of confidence.

PROJECTS and QUESTIONS

1. Investigate and report the contributions of the Greeks—especially Plato and Aristotle—to the development of modern science. Pay particular attention to their system of classification and the syllogism.
2. List ten false beliefs that were widely accepted, even by the intelligentia at one time or another in our cultural development. Trace the developments that led to their eventual rejection.
3. Report on the development of the periodic table in chemistry—or a similar achievement of scientific interest—as an example of the series of successive approximations to the truth characteristic of the refinement of science.

SELECTED REFERENCES

CAJORI, FLORIAN. "Baconian Methods of Scientific Research," *Scientific Monthly,* 20 (January 1925): 85–91.

CAMERON, D. EWEN. "The Current Transition in the Conception of Science," *Science,* 107 (May 28, 1948): 553–8.

COHEN, MORRIS R. and NAGEL, ERNEST. *An Introduction to Logic and Scientific Method.* New York: Harcourt, Brace, 1934.

CONANT, JAMES B. "The Role of Science in our Unique Society," *Science,* 107 (January 1948): 77–83.

CONANT, JAMES B. *Modern Science and Modern Man*. New York: Columbia University Press, 1952.

DINGLE, HERBERT (ed.). *A Century of Science, 1851–1951*. New York: Roy Publishing Co., 1951.

FAWCETT, HAROLD P. *The Nature of Proof*. 13th Yearbook; National Council, Teachers of Mathematics. New York: Teachers College, Columbia University, 1938.

GRUNBAUM, ADOLF. "Causality and the Science of Human Behavior," *American Scientist*, 40 (October 1952): 665–76.

HEMPEL, CARL G. *Fundamentals of Concept Formation in Empirical Science*. International Encyclopedia of Unified Science, Vol. 2, No. 7. Chicago: University of Chicago Press, 1952.

KANTOR, J. R. *The Logic of Modern Science*. Bloomington: Principia Press, 1953.

KAUFMAN, FELIX. "The Nature of Scientific Method," *Social Research*, 12 (November 1954): 464–80.

KELLEY, TRUMAN L. *Scientific Method: Its Function in Research and in Education*. New York: Macmillan, 1932.

LEWIN, KURT. "The Conflict between Aristotelian and Galilean Modes of Thought in Contemporary Psychology," *Journal of General Psychology*, 5 (April 1931): 141–77.

MORRIS, CHARLES W. "Scientific Empiricism," in Neurath, Otto, *et al.* *International Encyclopedia of Unified Science*. Vol. 1, No. 1. Chicago: University of Chicago Press, 1938.

NORTHROP, F. STUART C. *The Logic of the Sciences and the Humanities*. New York: Macmillan, 1947.

SARTON, GEORGE. *A Guide to the History of Science*. Waltham: Chronica Botanica, 1952.

SCATES, DOUGLAS E. "The Parallel Roles of Physical and Social Science," *Scientific Monthly*, 64 (January 1947): 14–20.

THOMSON, GEORGE. "The Two Aspects of Science," *Science*, 132 (October 4, 1960): 996–1000.

WIGHTMAN, WILLIAM P. D. *The Growth of Scientific Ideas*. New Haven: Yale University Press, 1953.

There is not a single phase of the educational process that a mature science of behavior could not render more effective

ROBERT M. W. TRAVERS

3 The Nature of Science

Purpose of Science

The basic purpose of science is the accumulation and clarification of experience, and the systematization of such experience into a relatively small number of broad general laws and principles governing the specific categories into which phenomena can be classified. It is not merely a matter of cataloging one's experiences or of describing their nature and characteristics in detail, but rather one of discovering or of establishing a structural system into which phenomena can be ordered and on the basis of which they can be predicted and, eventually, controlled. In the early stages of science, the task is to gather, define, and catalog experiences in order to obtain an understanding of their interrelationships. In the later stages of science, the task is to reduce to a minimum the number of laws necessary to express these relationships.

Interpreted broadly, the scientific method constitutes the most adequate approach to the discovery of truth, and certainly it has demonstrated its worth, particularly in the physical sciences. On the other hand, the limitations of the scientific method must be clearly understood. For instance, science cannot deal directly with values. It can define some of the issues involved in making value-judgments, but the judgments

themselves are outside the scope of science. Even the interpretation of the results of research is outside of the realm of science. And, while science attempts to minimize errors and guarantee valid results, errors have been made by persons using the scientific method. Usually, however, the errors were in the use of the method rather than in the method itself. It must also be noted that the scientific method does not lead to truth directly, but proceeds through a series of successive approximations to the isolation of more and more precise relationships and to a more and more adequate formulation of these relationships.

Science as a Method of Discovery

Science can be defined both as an organized body of knowledge and as a method and system of deriving truth. The latter is certainly the more crucial aspect, since it not only permits the discovery of knowledge but also affords us the means for accelerated scientific progress. The scientific method can be delineated into a number of steps, the exact formulation of which varies somewhat from writer to writer. The general pattern is: certain phenomena are observed; a problem situation develops and is noted and clarified; crude relationships are tentatively identified and elaborated; a more or less formal hypothesis is derived; a design is developed to test the hypothesis; the hypothesis is verified or refuted; the results are subjected to further tests and refinement; and finally, the conclusions are integrated with the previous concepts of science. The process involves such subsidiary steps as the review of related experience, the manipulation of factors, the measurement of quantities, the scaling of variables, and the analysis and interpretation of data. It must also be realized that the steps, while usually listed in a one-two-three fashion, rarely occur in that sequence, since the effective use of the scientific method does not allow for this sort of rigidity.

Obviously, not all attempts at discovering truth comply with the specific formulation presented above. Some scientists suggest that we ought to think of scientific *methods* rather than of *the* scientific method; they feel that using the term in the singular implies that there is only one right way to attack a problem, and that it leads people to confuse the scientific method with experimentation which is only one form of the

scientific method. Other scientists do not agree: they realize that though there are different ways of treating different problems, the need is for many different scientific *techniques* subsumed under the scientific method.

The difficulty is, obviously, one of definition of the terms, *science* and *scientific method.* If we rigidly hold to the scientific method described above as the criterion for inclusion or exclusion in the select club of science, we would have to exclude Einstein's contributions because they were almost exclusively organizational and deductive rather than experimental. And, unless we extend the concept of verification to include proof by observation, we also exclude such sciences as astronomy and geology in which manipulation of variables is relatively impossible. We also eliminate historical studies and many others in which control is relatively limited. The scientific method must not be interpreted so narrowly that it excludes all approaches except those in which the investigator actually manipulates—either physically or statistically—the conditions of his "experiment," and "causes" the occurrence of the events he wishes to observe.

We must also realize that the great scientific discoveries of history, as well as those of more recent times, were generally not achieved through close adherence to the formal steps of the scientific method as taught in our schools where, according to Kruglak, the spirit of scientific inquiry involved in "experiments" too frequently degenerates into collecting the same data that millions of other bored students have collected for years.[1] Furthermore, it must be noted that the ultimate in scientific progress and excellence is obtained, not through experimentation as such, but through the organization and systematization of scientific thought. In the more advanced stages of science, experimentation is essentially restricted to the role of confirming the outcomes which logical deduction has led the scientist to expect.

The Products of Science

Of great significance in our understanding of modern science is the change that has taken place in the way we view its products. Until the turn of the century, scientific "facts" were immutable. Once obtained, they were considered laws of

[1] Haym Kruglak, "The Delusion of the Scientific Method," *American Journal of Physics,* 17 (January, 1949): 23–9.

nature that, barring errors in derivation, had everlasting validity. Our newer concept of science as a matter of successive approximation to the truth casts the facts of science in the role of working hypotheses which have "not a validity, but a utility,"[2] in the sense that they are effective in expanding our control of events, and are valid only until a better working hypothesis is found. The so-called "laws of nature" are now considered simply a representation of our way of conceptualizing data, with everything relative and conditional.

The end-products of empirical science are the laws and generalization which place a number of relatively isolated relationships—discovered through the various techniques of science—into a unitary conceptual framework.

Science is generally oriented toward the derivation of laws which are of a nomothetic nature—that is, general laws which apply to all individuals of a given class or a given set. For example, bright children learn faster than dull children. Such laws are derived statistically and apply statistically—that is, the relationship is one of probability, and though they are stated as absolutes, it is more correct to state that bright children *tend* to learn . . . , or, *in general,* bright children . . . .

Such generalizations are not exception-free. They express a useful relationship, but they generally are of limited use in the individual case, since they are derived on the basis of an average set of conditions which no one can duplicate. To the extent that the individual complies with the conditions postulated in the law only in the general sense, the law applies to him only in a general way and, therefore, can be interpreted only on the basis of probability. This is perhaps more obvious in the social sciences where, for example, predicting which student applicant will be successful and which will be unsuccessful is always a precarious undertaking. The predicament, however, is not restricted to the social sciences: it pertains equally well to the bombardment of alpha particles where it is impossible to tell which particle will fly free and which direction it will take.

The degree to which a generalization can be applied to the individual case depends to a great degree on the ex-

[2] D. Ewen Cameron, "The Current Transition in the Conception of Science," *Science,* 107 (May 28, 1948): 553–8.

tent of the clear-cut, nonoverlapping distinctions stated in the conclusions. Thus, the generalization that adults are taller than children can be applied in the individual case with a high degree of probability, while the generalization that executive positions are occupied by persons who are taller than average has only a slightly greater than 50–50 probability of successful prediction in the individual case. The number of conditions included in the statement of the law also affects it applicability: the more of the relevant conditions in the antecedent of the law which are specified, the greater the probability of its validity in cases meeting those specifications. In fact, a law would be exception-free if its statement covered all antecedent conditions involved. However, then there would be no "new" cases and the law would be useless. Thus the more conditions we specify in the statement of a law, the more precise it becomes (under the conditions of its statement), but also the more restricted it is in its application—and the more useless it is as a tool of prediction.

Of particular interest to social scientists are the laws known as *idiographic*—that is, laws pertaining to the individual case. The clinical psychologist, for example, would rely on idiographic laws in predicting the behavior of his client on the basis of the specific characteristics that make this client like and yet different from other counselees. Each child in the classroom also is a unique individual whose behavior is governed by idiographic as well as nomothetic laws.

THEORY AND SCIENTIFIC PROGRESS

The Need for Organization

In its early stages of development, the major concern of science is the accumulation and refinement of experience and the discovery of functional relationships between phenomena through the application of methods of observation of various levels of refinement. As long as the relationships so derived remain isolated, however, they are of limited value except in the solution of a problem identical to those which led to their discovery. To be useful, knowledge must be organized, and the primary responsibility of a science is to develop a system of organization which will make the facts, as they are accumu-

lated, meaningful from the standpoint of their ultimate purpose.

Only when isolated facts are placed in perspective, as a result of being integrated in some conceptual scheme which promotes a greater understanding of their nature and significance, do we approach a science. Thus, Conant suggests that unless progress is made in reducing the degree of empiricism in an area, the rate of advance in that area will be relatively slow and highly capricious.[3] Similarly, McConnell points out that the development of a science depends as much on the continuous formulation and revision of theory as it does on investigation and experimentation.[4]

Science is never satisfied with isolated facts but is committed to a continuous process of ever-expanding clarification and systematization of its findings. Through continuous observation and experimentation, it attempts to evaluate the adequacy of previous generalizations and to isolate the conditions under which these previous generalizations can be expected to hold. Thus, in sequence, from simple experiences come simple hypotheses, which lead to further experience, further clarification, and more sophisticated hypotheses. As these hypotheses are substantiated, they become laws and principles which, as a result of being mirrored against further facts, hypotheses, and laws, become integrated theories. The ultimate goal is a systematization not only of facts into laws but of laws into ever-expanding conceptual schemes of science and an ever-smaller number of broad principles and theories. Thus the method of science is essentially one of a back-and-forth movement—from facts, to hypotheses, to laws, and back to facts as the basis for the testing and refinement of more adequate hypotheses; thus leading to the derivation of more general and comprehensive principles and theories. It must also be noted that generally the steps are rather small; progress in science is made by the slow accumulation of small steps and, frequently, the correction of missteps.[5]

[3] James B. Conant, "The Role of Science in Our Unique Society," *Science*, 107 (January, 1948) : 77–83.

[4] Thomas R. McConnell, *The Psychology of Learning*. 41st Yearbook, National Society for the Study of Education, Pt. 2 (Chicago: University of Chicago Press, 1942) , p. 8.

[5] Benton J. Underwood, *Psychological Research* (New York: Appleton-Century-Crofts, 1957) , p. 9.

Purpose of Theory

The purposes to be served by theory in the development of science have been implied repeatedly in the previous sections. They can be summarized as follows:

1. Theory synthesizes isolated bits of empirical data into a broader conceptual scheme of wider applicability and predictability. It permits deeper understanding of data and translates empirical findings into more readily understood, more readily retained, and more readily adaptable form. The theory of oxidation, for instance, places many of the chemical reactions common to everyday life in focus. Theory provides facts with meaning and significance by clarifying them and by placing them in perspective with one another and with preexisting theories. It defines the problem and its setting, and determines the relevancy of the facts that have been obtained.
2. Theory permits the prediction of the occurrence of phenomena and enables the investigator to postulate and, eventually, to discover hitherto unknown and unobserved phenomena. At the time the periodic table was being completed, for instance, certain gaps were noted in the sequence of the elements. Since according to theory, there should have been no gaps, scientists were spurred to look for the missing elements. In time these were found, probably much earlier than they would have been had their presence not been anticipated by theory. This brings out the interesting point that generally facts are gathered first and then explained, but here the opposite is true. The facts were explained first and then discovered. There are other "facts," of course, that are still in the postulated stage and have yet to be verified; some of these may never be discovered and perhaps do not even exist.
3. Theory acts as a guide to discovering facts; it pinpoints crucial aspects to be investigated and crucial questions to be answered. By identifying areas in need of exploration, it stimulates research in areas that are lagging.
4. Theory is based on the assumption that detailed empirical findings are special cases of more general laws, and that progress cannot be made as long as observations are simply accumulated. Theories cannot develop without experimental facts any more than the discovery of experimental facts can proceed far on the basis of grossly inadequate or incorrect theories. For example, the progress of psychiatry as a science was bound to be limited as long as the insane were viewed as

possessed by a devil. Just as facts underlie theories, theories underlie facts—each raising the other on a spiral to evermore precise scientific formulations. Research and theory go hand in hand: theory guides and stimulates research while research tests and stimulates theory development, resulting in more adequate theories and better and clearer facts. This is again a statement of the successive approximations and redefinitions on the basis of which scientific progress is made. Facts derive their significance from the theoretical framework into which they fit, just as theories derive their acceptability from the extent to which they bring facts into clearer focus. This is well stated in the following quotations:

> . . . [T]here is a constant and intricate relationship between facts and theory. Facts without theory or theory without facts lack significance. Facts take their significance from the theories which define, classify and predict them. Theories possess significance when they are built upon, classified, and tested by facts. Thus, the growth of science is dependent upon the accumulation of facts and the formulation of new or broader theories.[6]
>
> To conduct research without theoretical interpretations or to theorize without research is to ignore the essential functions of theory as a tool for achieving economy of thought.[7]

This is particularly true since, in order to achieve control and replicable results, research must confine its efforts to seeking answers to problems that have been delineated and controlled to the point that the outcomes are highly fragmented and isolated. There is a need, therefore, to organize the tiny, rigorously defined bits of knowledge into a more realistically meaningful setting. This is precisely the function of theory.

The Modern Acceptance of Theory

The value of theory is readily acknowledged by such notable scientists as Einstein and Conant and attested to by the fact that many of the world's leading scientists, for example, Newton, Poincaré, Whitehead, and Einstein, were philosophers of science rather than experimenters. In fact, it is the feeling of many scientists that we are limited not so much by the inade-

[6] Deobold B. Van Dalen, "Relationship of Fact and Theory in Research," *Educational Administration and Supervision*, 45 (September, 1959): 271–4.
[7] Claire Selltiz, *et al., Research Methods in Social Relations*. (New York: Holt, Rinehart, and Winston, 1959), p 199.

quacies of our techniques of research as by the inadequacy of our theoretical framework. There are, however, others who are more reserved in their support of theory. A relatively strong stand against theory is taken by Skinner[8] who questions the need for theories of learning on the grounds that they are not essential to the designing of significant experiments, and further that they actually retard the growth of science by diverting research time and effort from research proper and by limiting research to predetermined areas, while discouraging the investigation of those aspects which are not in line with current theory.

The layman, and even the professional person, frequently displays a lack of appreciation of the complementary nature of theory and practice. There is a tendency for the practical man to look on theory as something impractical and idealistic. It is true, of course, that many theories have not been well formulated and that many are based on speculation rather than on scientific fact. This is understandable: in the early stages of his work, the theorist must be blind to exceptions; otherwise, he would never be able to get started. He must bypass certain problems until he has gathered enough facts to resolve them. It is also true that, in the social sciences especially, many theories are relatively lacking in validity and scope, as well as in practicality, simply because they are not sufficiently advanced. It must not be assumed, however, that theory consists of blind speculation. On the contrary, a theory is an attempt at synthesizing and integrating empirical data for maximum clarification and unification—and certainly nothing is as practical as a sound theory.

Actually everyone has a number of personal theories based on postulates and assumptions of varying degrees of adequacy and truth from which he makes deductions of various degrees of cruciality and, of course, of accuracy. The principal, for instance, has many theories about education. These are based partially on personal experience, partially on his reading of relevant literature, and partially on his personal philosophy. But he looks upon these theories as practical facts, and he bases decisions on them as if they were truth. Rarely, if ever,

[8] Burrhus F. Skinner, "Are Theories of Learning Necessary?" *Psychological Review* 57 (July, 1950) : 193–216.

does he subject his theories to a test through a valid experiment of their logical deductions.

Probably the most fundamental determinant of the potential contributions of theory to a given science is the state of development of that science. It seems logical that in the early development of any science, the empirical approach—that is, the accumulation of data—must be paramount. In the later stages, on the other hand, theory is likely to become progressively more vital to its further growth. The question then is not one of whether one believes in the crucial role of theory in the advance of science, but whether a given science is ready for emphasis on theory. This is the position taken by Traxler[9] and also by MacKinnon[10] who questions the practice of requiring a doctoral candidate to take a definite theoretical stand in his doctoral dissertation, when the various members of his committee would probably not agree on a single theoretical position themselves.

It is fully recognized that premature subscription to a theory may blind the scientist to the correct solution in much the same way as the flat-world concept or the veneration of Aristotle may have delayed science for centuries. Furthermore, the fact that a discovery which is compatible with a theory is easier to accept than one which is not may lead to the perpetuation of false theories supported by prejudged or partial evidence. Premature attempts to reach a formalized theoretical position are likely to lead to the investigation of the more trivial aspects of science, simply because they are easier to conceptualize and to test. The more trivial aspects frequently lend themselves more readily to mathematical formalism, for example, with the resulting neglect of the more significant but more theoretically complex aspects. Another difficulty which arises in the social sciences is that the formulation of a theory immediately leads to its contamination, in the sense that people are affected by knowledge of the theory and, reacting to this knowledge, interfere with its realization. It also must be realized that a theory does not provide answers; it may stimulate and

9 Arthur E. Traxler, "Some Comments on Educational Research in this Century," *Journal of Educational Research*, 47 (January 1954) : 359–66.

10 Donald W. MacKinnon, "Fact and Fancy in Personality Research," *American Psychologist*, 8 (April 1953) : 138–46.

direct research, but it is through significant research and not through theory that significant answers are obtained.

The emphasis to be placed on theory in the development of a given science at a particular stage of development is difficult to establish since it revolves around the question of when enough is enough. Undoubtedly, a science is always ready for theory—at its own level, of course—but ready, nevertheless. It is not fruitful to concentrate on the accumulation of data without some (perhaps vague) idea of what is sought. The accumulation of data and the organization of these data into theoretical structure must go hand in hand, and any lag in one is bound to cause a corresponding lag in the other.

Theory as a Point of Reference

All sciences make use of deduction to some degree. In fact, though in the beginning a given science must concentrate on the accumulation of evidence and the inductive development of tentative hypotheses, in its later stages the relative ratio of induction to deduction leans more toward deduction. Much of the effort of the physicist, for example, is devoted to the mathematical manipulation of previously derived formulas of the relationship among phenomena. In fact, in the more advanced sciences, scientists place their maximum concern on the development of theory on the basis of which empirical observations are to be guided and explained. A physicist, for instance, would look with some degree of suspicion on any result that he could not integrate with previously established theory. This is not to imply that physics is fully explored and complete, but rather that it is sufficiently stable and integrated that the next improvements are likely to be small changes, and, furthermore, changes that are compatible with present views. Thus, Einstein's theory of relativity gained early acceptance because it was a better explanation, rather than a refutation of what was already known about gravity and electrical theories and reconciled some of the contradictions of earlier theories.

When exceptions to theories arise, they must be integrated into more adequate theories. A classic example of the resolution of such exceptions can be found in the case of the Dulong–Petit Law, which stated that the specific heat of a solid

element multiplied by its atomic weight is a constant (approximately 2 calories per degree). At first, this was simply an empirical observation, with no scientific explanation. Two notable exceptions—carbon and silicon—were known, however. Later, with the development of the Quantum Theory, not only was the relationship explained, but the exceptions themselves were also explained as special cases.

Data in agreement with theory do not prove a theory, but merely support it; but a single item of negative evidence is logically sufficient for its rejection or its modification. However, scientists have operated on the premise that a theory is not so much true or false as it is useful or useless, and that an inadequate theory is probably better than no theory at all. Generally, an exception calls for some adjustment in a theory rather than its complete scrapping. As Conant points out, "A conceptual scheme is never discarded merely because of a few stubborn facts with which it cannot be reconciled; a conceptual scheme is either modified or replaced by a better one, never abandoned with nothing left to take its place."[11]

Exceptions to a theory actually serve a very useful purpose in promoting crucial research and, eventually, in improving the theory. This is not to imply that a theory, once established, except for minor refinements, will stand forever. Certainly the phlogiston theory had to be abandoned completely in favor of the oxidation theory, but in the usual case a refinement or perhaps an extension of the theory can incorporate new evidence. Actually, a theory is rarely, if ever, complete: as new facts appear—and in most theories new facts appear endlessly—theories have to undergo some modification. Rarely, however, is there need for a completely new conceptual scheme.

A particularly important change that probably needs to be made in the future is the unification of the theories used in the different disciplines to explain their own particular data. Each field has developed a considerable amount of specialized knowledge, and a major task of science seems to be that of building bridges from one discipline to another in order to integrate this specialized knowledge into a single conceptual structure. This assumes the unity of science—that is, it assumes

[11] James B. Conant, *Science and Common Sense* (New Haven: Yale University Press, 1951), p. 170.

not only that the universe is subject to law and order, but also that various components are subject to the same set of principles of law and order. This seems reasonable—at least more reasonable than to assume that the principles that govern the aspects of one discipline are independent of or in conflict with those of another. Some basis for such unification already exists in the common allegiance of the various disciplines to the scientific method. There is also considerable communality among the basic concepts fundamental to the various fields. The clinical psychologist facing his client, for example, may well be reminded of Newton's third law of motion—that is, that an object will continue at rest if at rest, or in motion if in motion, until it is affected by a force. Nor are such concepts as *field forces* and *valence* used by Gestalt psychologists too remote from their counterparts in the physical sciences. In recent years, there has been a great deal of rapprochement in the principles and laws of the sciences of the physical and social order, and theoretical unity has become a feature of the more advanced sciences.

Psychologists have developed a number of "schools of psychology" which are not only in disagreement with one another, but which have yet to develop a wholly consistent and satisfactory explanation of all psychological phenomena. There is, of course, a major need for the unification of the various theories, and a start has been made in this direction. McConnell,[12] for example, attempted to identify the crucial points of difference among the various theories of learning and the issues which need to be resolved in their reconciliation. Education is almost completely lacking in a consistent theory. Whereas, in the days of Thorndike, much of the work of education was co-ordinated on the basis of connectionism, present-day educators tend to subscribe to an eclectic and, at times, self-contradictory approach. A teacher may, for example, talk of the whole child while drilling arithmetic combinations!

It is not inconceivable that some day a single theoretical system will be used to explain the behavior of molecules, of animals, and of people. Even today, some of the features of the field theory, for example, with its concept of field forces, apply to people in their environment as well as to electrons in their various shells and to heavenly bodies in orbit. Hartmann

[12] Thomas R. McConnell, *op. cit.*, Ch. 7.

points out that field theories claim such scientists as Whitehead, Planck, and Einstein in the physical sciences; Cannon, Lashley, Woodger in physiology; Wertheimer, Kohler, and Koffka in Gestalt psychology; and Lewin in topological psychology.[13] Thus, field theory in a sense, is more a theory of science than it is simply a school of psychology, and it might conceivably contain the seed for a more complete unification of scientific phenomena.

Characteristics of a Good Theory

The extent to which a given theory can be expected to fulfill its purposes is dependent on the extent to which it meets certain basic criteria. Among these are:

1. A theoretical system must permit interpretations and deductions which can be tested empirically—that is, it must provide the means for its own interpretation and verification. Much of the work of Freud, for instance, does not provide testable deductions and is, therefore, a matter of speculation rather than of scientific theory.
2. Theory must be compatible both with observation and with previously validated theories. It must be grounded in empirical data which have been checked and verified and must rest on sound postulates and hypotheses. The better the theory, the more adequately it can explain the phenomena under consideration, and the more facts it can incorporate in a meaningful structure of ever-greater generalizability. A good theory is one that has as wide an applicability as the present state of knowledge will permit.
3. Theories must be stated in simple terms; that theory is best which explains the most in the simplest form. This is the *law of parsimony*. A theory must explain the data adequately and yet must not be so comprehensive and detailed as to be unwieldly. On the other hand, it must not overlook variables simply because they are difficult to appraise. A theory must be stated precisely and clearly, if it is to serve as an adequate guide to research.
4. Scientific theories must be based on empirical facts and relationships. The mere accumulation of empirical data, however,

[13] George W. Hartmann, "The Field Theory of Learning and its Educational Consequences," *The Psychology of Learning*. 41st Yearbook, National Society for the Study of Education, Pt. 2 (Chicago: University of Chicago Press, 1942), p. 173.

constitutes neither theory nor science, until the data have been organized into general principles that permit the interpretation of particular phenomena on the basis of the operation of more fundamental underlying factors. As Travers points out, these can range from highly formalized theories involving fully developed mathematical relationships to those that are yet informal, such as those in education.[14]

The more highly developed a science becomes, the more likely theories are to shift from constructs involving events that can be experienced directly to those that have to be inferred. theories must not be defined in hypothetical constructs that are circular. We cannot, for example, postulate that a child does not want to work because he is lazy, since laziness cannot be defined except as not wanting to work. Nor can we base a theory on concepts that have not been shown to exist. It also must be remembered that hypothetical constructs are simply aids to explanation; they must not be used as if they existed in reality.

The Role of the Theorist

In view of the complementary nature of research and theory, both need to be pursued with equal zeal if science, as a unitary discipline, is to progress. Researching and theorizing go hand in hand, and it is generally desirable to begin the report of an investigation by fitting the study into the framework of existing theory, and to end it by pointing out the implications of the findings and conclusions according to their theoretical as well as their practical significance. Thus the scientist is both an investigator and a theorist.

It does not follow, however, that a scientist is equally skilled in these two essential but somewhat independent aspects of science. Without subscribing to the stereotype of the scientist as a man of solitude, few words, and a very specialized and restricted background, who is not too well suited for theory development, we need to recognize that it may be difficult for a scientist to perform both as investigator and as theorist. On the other hand, there are scientists with good general background and insight into a particular field and special ability at organization and expression, who could make a significant contribution to science by developing its theoretical

[14] Robert M. W. Travers, *An Introduction to Educational Research* (New York: Macmillan, 1958), Ch. 2.

framework, leaving to others the task of delving more deeply and intensively into some of its more precise and restricted aspects.

RESEARCH AS AN ASPECT OF SCIENCE

Nature of Research

The confusion which surrounds the meaning of *research* is even greater than that which surrounds the term *science*. First, it must be repeated that at no time in history did men *begin* to do research; even primitive man attempted to seek truth from his environment and, in a sense, was doing research. The term, as it is used today, however, is restricted to the more systematic and formal search for orderliness among phenomena. Research may be defined as the systematic, objective, and accurate search for the solution to a well-defined problem. Best refers to research as "the formal, systematic, extensive process of carrying on the scientific method of analysis."[15] He points out that while one can be scientific without doing research, one cannot do research without being scientific.

A somewhat narrower definition restricts research to the fifth step of the scientific method—that is, to the testing of the hypothesis—and places the remaining steps of the scientific method more or less outside the realm of true research. Most people would reject this narrow definition, for it takes research out of the overall context of science and makes it meaningless and unprofitable. Such a definition raises the questions: Was Einstein's derivation of the atomic bomb research? Or Dewey's formulation of the steps of critical thinking? A question could also be raised as to whether the testing of the hypothesis is the crucial aspect of research: What if Einstein knew beforehand that the atomic bomb could be devised, so that the testing of the bomb was simply a formality for the purpose of working out a few technical details?

The worth of a scholarly enterprise is not gauged exclusively by its compliance with the criteria of the scientific method in its narrowest sense. With the increasing urgency for the synthesis of research findings, for instance, scholarly

[15] John W. Best, *Research in Education* (Englewood Cliffs: Prentice-Hall, 1959), p. 6.

writing (though not research in the usual sense) may constitute a greater contribution to science than does research on a trivial problem no matter how adequately it meets the criteria of science. Might it not be more profitable to define research on the basis of its contribution to the attainment of truth—either through discovering heretofore unknown relationships among phenomena or through establishing a greater degree of orderliness among what is already known? It is all a matter of semantics; there is, however, no point in defining research so narrowly as to rob it of its significance.

Pure and Applied Research

Progress in science is best promoted by proper emphasis on the dual processes of deriving knowledge and of organizing such knowledge into a theoretical structure. The scientist must devote himself with equal vigor to the pursuit of both, and both must be held in equal honor. In practice, however, it is frequently difficult to maintain such a balance.

Man is always faced with problems, some immediate and some remote. He hopes that eventually most of his problems will be solved. In the meantime he has to cope with the present as well as the future, with the present frequently having priority. It seems logical that he will accomplish more in solving his problems—both remote and immediate—by developing the required theory and by deducing the solution to his immediate problems from the general theory. Since some of the problems he faces are here and now, however, pressing him for an immediate solution, the necessary solution may be obtained more quickly by seeking it directly rather than indirectly through the development of the required theoretical framework. The present-day emphasis on operations research in industry and action research in education are both oriented to the solution of problems of the immediate situation at the empirical level. Such research is frequently performed at a low level of scientific sophistication, and, at best, is of limited generalizability, though it may provide hypotheses for more careful research at a later date.

Both pure and applied research are oriented toward the discovery of truth, and both are practical in the sense that they lead to the solution of man's problems. Research is research

even though it has no immediate, or even forseeable, practicality. Furthermore, all research probably will be useful and practical eventually, no matter how pure and removed from practicality it is at the moment. Yet, from the standpoint of the directness with which the solution to the immediate problem is sought, a distinction can be made between pure research, which is interested in the theoretical aspects of science and only indirectly in the practical application which these findings may have, and applied research which has exactly the opposite orientation. The major question seems to be whether as great a contribution to the development of science and the welfare of mankind can be made by concentrating on pure research as is made by devoting the same amount of time and energy to the solution of immediate practical problems. Most scientists would reply that pure research contributes more to the long-range advancement in science.

The practitioner, faced with problems here and now, cannot wait. Furthermore, he has discovered that theory is not always right, or, more specifically, that while a solution may be right for the conditions under which it was derived, it may not apply to his particular case so that he will still have to solve his own problem—perhaps with improved insight, but nevertheless on his own. He is frequently impatient with the artificial nature of the theoretician's problems and his neglect of real problems. Periodical reactions set in against theory. For example, the Depression saw a movement toward the M. Ed. and the Ed. D. as "practical" degrees in contrast to the M.A. and Ph.D. with their greater emphasis on theoretical considerations. To the extent to which any real difference exists between these degrees, there is an implication that the cause of education is best served by emphasis on the solution of practical problems rather than on the derivation of theoretical structure. Action research is another indication of the educator's impatience with theory.

It must be recognized that, though systematic theory has contributed directly to the solution of practical problems, the contribution has not been entirely one-sided. Applied research in the solution of immediate problems has also contributed to the clarification of theory through the suggestion of valuable hypotheses and general insights that can be incorporated into a

more sophisticated attempt at pure research. In fact, a successful attack on theoretical concepts often must await the development of a certain degree of lower level applied research. The scientific benefits that might accrue from practical research, however, are frequently lost through failure to relate empirical findings to their theoretical implications. Too often, all that is derived from such studies is the solution to an immediate problem—plus a vague set of rules of thumb that are of doubtful, limited, and restricted validity.

THE RELATION OF SCIENCE AND PHILOSOPHY

The conflict between empiricism and theory finds a parallel in the conflict between the roles of science and philosophy. The former distinction is between the accumulation of data relative to natural phenomena and the integration and unification of the relationships obtained into an underlying conceptual structure. The present distinction is between the proper understanding of the empirical and theoretical nature of phenomena and the interpretation of such phenomena according to human goals and purposes.[16]

Although science and philosophy exist in an interdependent and complementary role, this has apparently not always been apparent in the behavior of either the philosopher or the scientist. The philosopher seems to feel that the important things life—human goals and values—are not subject to scientific determination. He tends to look down on the scientist whose concern is often materialistic and who frequently attempts to discover truth through consensus, statistical manipulation, and the concept of probability. The scientist, on the other hand, seems to feel that science has led to our progress and our material welfare, and that philosophers are dreamers whose concern with values frequently takes the form of speculative and intuitive deductions boosted to the level of dogma through emphatic pronouncement. This opinion is frequently shared by the layman, who thinks that "what research says" is accurate and dependable, while "what philosophy says" is speculative and generally undependable.

There is need both for the scientist to understand the

[16] The term *philosophy* is frequently confused with the concept of theory development, discussed in the previous section.

philosopher and for the philosopher to understand the scientist: they are pursuing the same goal. Their methods are different—in science, accuracy is largely a function of control, replication, and randomness; in philosophy, dependability is based on the accuracy of the definition of the problem, the recognition of the basic assumptions, and the accuracy of the logical processes. Yet there is no need for conflict between the two, and many of our great scientists have been able to combine the two functions into one.

From the standpoint of their function, science answers the question "what?"; philosophy answers the question "to what end?" Science is concerned with the discovery of knowledge; it can tell what is and why. But this is only a means to an end; philosophy begins where science leaves off, and is concerned with the use of this knowledge. The task of science is to determine the most efficient way of attaining a certain goal; whether that goal is worthy of attainment is a philosophical consideration. Thus, philosophy is concerned with the ultimate ends toward which research needs to be oriented. Philosophy provides the framework within which a problem can and does exist. Science works with the means; it can improve the efficiency of the process, but it cannot resolve the question as to the desirability of the end.

Science is efficient but amoral and can work as effectively toward the attainment of evil goals as it can toward the promotion of desirable goals. It can provide the most effective means for promoting competitive behavior or co-operative behavior, just as it can be used in concentration-camp experiments with human beings or in the cure of cancer. Science can provide the knowledge on which value-judgments can be based; it can provide information about the effects of various courses of action, and thus provide a perspective from which the desirability of each can be seen in clearer focus—but it cannot deal directly with the values themselves. But neither can philosophy make value-judgments without considering scientific foundations; any attempt to do so is bound to result in poor judgment. In the words of Freeman "Bad science is not cured by good philosophy, nor can good philosophy arise from bad facts."[17] Thus, a philosophical decision to orient our

[17] Frank N. Freeman, "The Contributions of Science to Education," *School and Society,* 30 (July 27, 1929): 107–12.

schools towards progressivism or traditionalism cannot be made without considering the scientific evidence regarding the likely outcomes of the two approaches. In the same way, the decision to use the atomic bomb is outside the particular province of science, but science can provide the facts that need to be considered in deciding whether or not to use it.

Science and philosophy play complementary roles and every problem has both scientific and philosophical components. Science derives knowledge. Philosophy determines the ends which this knowledge is to serve in fostering the major goals of the social order. It helps to define and clarify the problem to be solved and the assumptions under which the conclusions derived from science are true. And, of course, it interprets what has been found with respect to the goals of society.

THE SCIENTIST

Status of the Modern Scientist

Since Hiroshima, and especially since *Sputnik* I, American society (if not world society) has become progressively more conscious of the crucial role of the scientist in the progress and the survival of mankind. Scientists—particularly nuclear physicists—have found themselves in such high esteem that their opinions, even on nonscientific issues, are sought and frequently are accepted unquestioned. In fact, as pointed out by Michels[18] scientists now have a voice far out of proportion to their numbers in shaping national and international thought and policy. Furthermore, their opinions, which make the headlines and which have such powerful political and sociological influence, frequently fall directly outside the area of competence of its author.

If he is to wield such influence, it is necessary for the scientist to appreciate the nature of his role. First, it must be recognized that science itself is amoral, and that the scientist *per se* has neither obligation nor responsibility. The scientist is, however, also a person—or, more specifically, a citizen of a country —which, at once, puts him under both obligation and responsi-

[18] Walter C. Michels, "Limits of the Scientist's Responsibility," *American Journal of Physics*, 16 (May, 1948) : 289–94.

bility. In fact, since prestige is invariably bought at the expense of greater responsibility, the scientist is faced with moral problems beyond those of the average citizen. Society has the right to expect him to contribute to its welfare and advancement in keeping with his talents—as it does all of its citizens—both by producing the means for such advancement and by providing whatever leadership his potentialities and status permit him to provide. He is expected to contribute toward its goals, regardless of what his personal views may be. For example, he has a right to object to the use of the atomic bomb, but he must do so as a citizen. He has no right to jeopardize its development through his lack of co-operation any more than a soldier has the right to sabotage plans for the attainment of a military objective. This position, of course, is not one of unanimous agreement; there are those who argue that the scientist is also a person who must live with his conscience, and, just as the conscientious objector can refuse to bear arms in the defense of his country, so the scientist should be free to withhold his services and discoveries if he fears their misuse.

The social scientist is not engaged in anything quite so spectacular as the development and use of the atomic bomb, but he too has definite responsibilities. He too has the obligation of conducting whatever research into social phenomena which his status, position, and competence permit, and further to make known his findings for the enlightenment and betterment of the social order. He has a special problem, however, in that not only do his findings affect people directly, but rarely are his "discoveries" and his interpretations ironclad. It is therefore imperative that the social scientist have both a sufficient understanding of the philosophical and sociological considerations underlying his problem and a thorough grasp of the problem itself and the limitations of his findings, so that he can see his conclusions in their proper perspective. He must be particularly careful to avoid misinterpretation in presenting his findings.

Having convinced himself of the action dictated by his findings, the scientist has the further responsibility of striving for the adoption of his viewpoint. Two cases present themselves here: 1. If his findings and interpretations are matters of unanimous agreement among his fellow-scientists, he can proceed as

a scientist to explain the scientific position and to urge action. 2. If, on the other hand, this is not a matter of complete scientific agreement, he can operate only as a private citizen expressing a personal opinion. He must, then, be careful not to abuse his status and his prestige as a scientist in order to promote controversial views. In all cases, he must remember that science must be the servant and not the master of man; it must never replace judgment but simply define the issues involved.

The scientist must, furthermore, be careful not to use his position in his field to promote personal views in a field in which he has no right to speak as an expert. A nuclear physicist violates ethics when he advertises his status as a scientist to enhance his opinions on educational practices, psychological testing, and so on. The scientist will determine by the way he discharges his responsibilities and the way his behavior complies with high ethical standards whether he deserves the prestige accorded to him by our present society.

There is also the opposite problem of the scientist becoming so scientifically objective in his views that he develops moral detachment and skepticism to the point of losing moral perspective. Although the scientist must repress subjectivity and personal feelings when acting in his capacity as a scientist, for him to carry a similar attitude into his social world represents an abuse of science, a sort of sterile intellectualism.

Of course, the scientist must not allow feelings of inadequacy in fields outside of his specialty to cause him to become a non-participant in society. He needs to realize that he has a responsibility as a citizen to take part and that, though his knowledge of social problems may be somewhat inadequate, his views are probably as adequate as those of many of the people who do take part. He cannot avoid his civic responsibilities for in a democracy the abdication of good men from active citizenship is an open invitation for the forces of evil and of incompetence to take over. Bowman[19] advocates an emphasis on courses in sociology and philosophy as the means of counteracting the development of such skepticism. It would seem that if one ceases to be a man, there is hardly any point in being a scientist.

[19] Claude C. Bowman, "Must the Social Sciences Foster Moral Skepticism?" *American Sociological Review*, 10 (December, 1945) : 709–15.

Professional Ethics

The necessity for professional ethics has been fully recognized by society, in general, and by professional groups, in particular. The American Psychological Association, for instance, has a code of ethics defining the responsibility of psychologists to the profession, to their clients, and to the sponsoring agency. Although most of the emphasis in such codes is on professional practice, ethical problems are bound to arise in connection with the conduct of research. There is, for instance, the problem of coding questionnaires in order to identify respondents when they have been "allowed to remain anonymous." Matters of ethics are also involved in the use of school children as experimental guinea pigs, particularly when such experiments interfere with the teacher's effectiveness in fulfilling his primary responsibility of teaching children. It would have to be assumed that any harm done is more than compensated for by the greater good that comes from the derivation of more effective methods.

Characteristics of the Scientist

Although many of the characteristics of the scientist can be inferred from the previous discussion, there is no standard "scientific personality" that characterizes all scientists, least of all the stereotype of the scientist as a non-social "intellectual" who seldom goes out of his laboratory. There is, of course, a basic core of such fundamental traits as intellectual integrity, professional responsibility, and scientific skepticism which motivate all scientists to a great degree. The list of such traits is so long and the degree to which they are involved so flexible, however, that there is an unlimited range of individual differences even among the top scientists in a given field.

Many writers have presented lists of the characteristics they considered typical of scientists, but these lists are so comprehensive that they merely include most of the desirable scholarly traits. A more meaningful approach is that of Shannon who investigated the personality characteristics of two hundred fifty world-renowned research workers. Among the traits he found common to this illustrious group, he lists in order: 1. enthusiasm and research zeal; 2. intelligence, adaptability,

resourcefulness, and versatility; 3. creativity, initiative, originality, ingenuity, and intuitiveness; 4. expertise and competence in their area of investigation; and 5. determination and drive.[20] Other specific traits frequently encountered in the literature describing scientists—many of which overlap those mentioned above—include intellectual curiosity, open-mindedness, freedom from bias, persistence, and thoroughness.

Generally the university atmosphere is considered most conducive to the maximum development and productivity of the scientist. While industry frequently ties the scientist to the task of providing desired products, the university generally imposes fewer restrictions on his freedom. The ready availability of stimulation and of consultation with colleagues, as well as the continuous challenge provided by students, especially at the graduate level, are significantly favorable factors which tend to be denied the man in the field. This is probably especially true of education. The superintendent of schools, for example, is generally too busy for his own good, and he often lacks the challenge of other experts who can help sharpen his thinking.

On the other hand, if the university is to capitalize on the creative talents of its faculty, it needs to provide for the exercise of these talents by keeping teaching and other responsibilities to a level where creative activities are possible. Education, in particular, seems to suffer from excessive teaching loads, and many professors grow old without having engaged in any professional activities other than those connected with meeting their classes and attending committee meetings and conventions. The problem deserves serious consideration if education is to derive the full benefits of the talents of its members.

SUMMARY

1. The basic purpose of science is the systematization of experience into a structural framework on the basis of which the significance of phenomena can be grasped.

2. The scientific method, interpreted broadly, constitutes the most systematic and generally the most adequate approach to the discovery of empirical truth. It generally encompasses a series of steps consisting of the selection and clarification of a problem,

[20] John R. Shannon, "Traits of Research Workers," *Journal of Educational Research,* 40 (March, 1947): 513–21.

the derivation and elaboration of a hypothesis, the collection of data and the testing of the hypothesis, and finally the generalization of the results. The scientific method cannot be equated with the application of so many steps in rigid sequence, however; to be effective, it must allow for considerable flexibility in its use.

3. Empirical laws are the expression of certain regularities existing among phenomena. They are best conceived as simply working hypotheses which enable us to grasp phenomena more adequately but whose validity is only tentative. Empirical laws can be idiographic or nomothetic.

4. While in the early stages of its development, science's major concern is with the derivation of empirical relationships among phenomena, empirical science is limited in usefulness. Its many generalizations must be structured into a meaningful conceptual framework. The ultimate goal of science is not only the systematization of facts into broad empirical laws and principles, but also the systematization of empirical laws into an ever-smaller number of theories explaining the basis for the relationships noted. The ultimate need is for the unification of the laws and theories of the various disciplines into a single overall scientific—empirical and theoretical—framework.

5. Theory permits a deeper understanding of the significance of phenomena, anticipates hitherto unknown relationships, and acts as a guide to meaningful research in productive areas. In practice, there must be a back-and-forth movement from the discovery of empirical facts and the structuring of these facts into a conceptual scheme and the orientation of research toward the discovery of further facts that will permit the derivation of more adequate theories. Although premature theoretical rigidity can lead research astray, there is a need in education for a greater appreciation of the complementary role of the empirical and theoretical phases of science.

6. A theory can never be proved; it can only be accepted if it provides an adequate explanation of empirical facts, or it can be rejected. In practice, however, a theory is not so much true or false as it is useful or useless, and theories, even though apparently false, at least in part, tend to last until modified or replaced by more adequate theories. Meanwhile, the very process of verifying a theory frequently serves a definite purpose in clarifying underlying concepts and in orienting research efforts in meaningful directions.

7. A good theory—just as a good hypothesis—must provide a more parsimonious explanation of the empirical facts discovered than any competing theory. It must especially be amenable to empirical validation.

8. Research has been defined as the systematic, objective, and accurate search for the solution of a well-defined problem. Any systematic and scholarly activity designed to promote the development

of education as a science can be considered educational research. Research may be classified as pure research and applied research. The latter is the more immediately practical, but the former generally makes the greater contribution to scientific progress.

9. Science and philosophy exist in inter-dependent and complementary roles. Science provides knowledge concerning the most efficient means of attaining certain goals; philosophy is concerned with the worth of these goals.

10. The scientist plays a crucial role in the welfare and progress of mankind. He needs to appreciate the special responsibilities that accompany the prestige which modern society has accorded him.

PROJECTS and QUESTIONS

1. a) List some of the laws, principles, and theories of interest to educators. What is their present status?
   b) What are some of the basic assumptions of modern educational thought and practice? What seems to be their validity?
2. a) Compare the scientific status of education with that of the physical and biological sciences on the basis of their principles, laws, and theories.
   b) On the basis of the above study clarify the meaning of the terms *law, principle,* and *theory.*
3. Read the biography of two or three of the world's great scientists. What are some of the characteristics that might have contributed to their greatness?

SELECTED REFERENCES

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. *Symposium on Basic Research.* Washington, D.C.: The Association, 1959.

AUSUBEL, DAVID P. *Theory and Problems of Child Development.* New York: Grune & Stratton, 1958.

BECK, SAMUEL J. "The Science of Personality: Nomothetic or Idiographic?" *Psychological Review,* 60 (1953) : 353–9.

BENJAMIN, ABRAM C. "Science and its Pre-Suppositions," *Scientific Monthly,* 73 (September 1951) : 150–3.

———. "Is there a Scientific Method?" *Journal of Higher Education,* 27 (May 1956) : 233–8.

BOWLBY, J., *et al.* "Symposium on the Contribution of Current Theories to an Understanding of Child Development," *British Journal of Medical Psychology,* 30 (1957) : 230–69.

BOWMAN, CLAUDE C. "Must the Social Sciences Foster Moral Skepticism?" *American Sociological Review,* 10 (December 1945) : 709–15.

BRONOWSKY, JACOB. *The Common Sense of Science.* Cambridge: Harvard University Press, 1953.

———. "Science and Human Values," *University Quarterly,* 10 (May 1956) : 247–59; 324–38; 11 (November 1956) : 26–42.

BUSWELL, GUY T. "Science and Social Philosophy: The Scientific Point of View in Education," *School and Society,* 69 (January 1, 1949) : 1–4.

CARTWRIGHT, DESMOND S. "Annotated Bibliography of Research and Theory Construction in Client-Centered Therapy," *Journal of Counseling Psychology,* 4 (Spring 1957) : 82–100.

DAVITZ, JOEL R. "Contributions of Research with Children to a Theory of Maladjustment," *Child Development,* 29 (March 1958) : 3–7.

DEUTSCH, MORTON. "A Theory of Cooperation and Competition," *Human Relations,* 2 (April 1949) : 129–52.

DUMAS, FRANK M. "Science and the Single Case," *Psychological Review,* 1 (1955) : 65–76.

EINSTEIN, ALBERT. *Out of My Later Years.* New York: Philosophical Library, 1950.

EINSTEIN, ALFRED and INFELD, LEOPOLD. *The Evolution of Physics.* New York: Simon & Schuster, 1938.

ESTES, WILLIAM K., *et al. Modern Learning Theory.* New York: Appleton-Century-Crofts, 1954.

FREEMAN, FRANK N. "The Contributions of Science of Education," *School and Society,* 30 (July 27, 1929) : 107–12.

FRANK, PHILLIP G. (ed.). *The Validation of Scientific Theories.* Boston: Beacon Press, 1957.

GARRETT, HENRY E. "A Developmental Theory of Intelligence," *American Psychologist,* 1 (September 1946) : 372–8.

GORDON, GARFORD G. "Better Theories Needed in Research," *Phi Delta Kappan,* 34 (March 1953) : 219–21.

GUILFORD, J. PAUL. "The Structure of Intellect," *Psychological Bulletin,* 53 (July 1956) : 267–93.

HELLFRITZCH, A. G. *Fundamentals of Concept Formation in Empirical Science.* International Encyclopedia of Unified Science, Vol. 2, No. 7. Chicago: University of Chicago Press, 1952.

HULL, CLARK L. *Mathematico-Deductive Theory of Rote Learning: A Study in Scientific Methodology.* New Haven: Yale University Press, 1940.

HULLFISH, H. G. "The Relation of Philosophy and Science in Education," *Journal of Educational Research,* 20 (October 1929) : 159–65.

IHDE, AARON J. "The Inevitability of Scientific Discovery," *Scientific Monthly,* 67 (December 1948) : 427–9.

JASPERS, KARL. "Is Science Evil?" *Commentary,* 9 (March 1950) : 299–300.

KANTOR, J. R. "Current Trends in Psychological Theory," *Psychological Bulletin*, 38 (January 1941): 29–65.

KILPATRICK, WILLIAM H. "A Defense of Philosophy in Education," *Harvard Teachers Record*, 1 (November 1931): 117–22.

LAFLEUR, LAURENCE J. "On the Nature of Scientific Law," *Scientific Monthly*, 74 (May 1952): 247–52.

McCONNELL, THOMAS R. *The Psychology of Learning*. 41st Yearbook, National Society for the Study of Education, Pt. 2. Chicago: University of Chicago Press, 1942.

McCONNELL, THOMAS R., *et al. The Conceptual Structure of Educational Research*. Supplementary Educational Monograph, No. 55. Chicago: University of Chicago Press, 1942.

MAIER, NORMAN R. F. and ELLEN, PAUL. "Integrative Value of Concepts in Frustration Theory," *Journal of Consulting Psychology*, 23 (June 1959): 195–206.

MARQUIS, DONALD G. "Research Planning at the Frontiers of Science," *American Psychologist*, 3 (October 1948): 43–8.

MARTIN, CLYDE V. "The Place of Philosophy in Educational Research," *Phi Delta Kappan*, 40 (October 1958): 39–40.

MARTZ, VELORUS. "Philosophy and Science," in Monroe, Walter S. (ed.). *Encyclopedia of Educational Research*. New York: Macmillan, 1941.

MERTON, ROBERT K. "The Bearing of Empirical Research on the Development of Social Theory," *American Sociological Review*, 13 (October 1948): 505–15.

MILLER, JAMES G. "Toward a General Theory for the Behavior Sciences," *American Psychologist*, 10 (August 1955): 513–31.

MOLES, OLIVER, *et al. A Selective Review of Research and Theories Concerning the Dynamics of Delinquency*. Ann Arbor: Institute for Social Research, 3rd ed.; New York: Macmillan, 1960.

MOWRER, ORVAL H. "Two-Factor Learning Theory: Summary and Comment," *Psychological Review*, 58 (September 1951): 350–4.

RATHS, LOUIS E. "Questions and Comments on Classroom Researches into the Needs Theory," *Journal of Educational Sociology*, 24 (March 1951): 426–32.

RESTLE, FRANK. "A Theory of Discrimination Learning," *Psychological Review*, 62 (January 1955): 11–9.

ROGERS, CARL R. "Toward a Theory of Creativity," *ETC, Review of General Semantics*, 11 (Summer 1954): 249–60.

———. "Client-Centered Therapy," *Journal of Counseling Psychology*, 3 (Summer 1956): 115–20.

ROHRER, JOHN H. "An Evaluation of College Personnel Work in Terms of Current Research on Interpersonal Relationships," *Educational and Psychological Measurement*, 9 (Fall 1949): 420–44.

RYANS, DAVID G. "Theory Development and the Study of Teacher

Behavior," *Journal of Educational Psychology,* 47 (December 1956) : 462–75.

SEARS, ROBERT R. "A Theoretical Framework for Personality and Social Behavior," *American Psychologist,* 6 (August 1951) : 476–82.

SIGEL, IRVING E. "Need for Conceptualization in Research on Child Development," *Child Development,* 27 (June 1956) : 241–52.

SKINNER, BURRHUS F. "Are Theories of Learning Necessary?" *Psychological Review,* 57 (July 1950) : 193–216.

———. "A Case History in Scientific Method," *American Psychologist,* 11 (May 1956) : 221–33.

SMITH, B. OTHANIEL. "Science of Education," in Monroe, Walter S. (ed.) . *Encyclopedia of Educational Research.* New York: Macmillan, 1950.

SPENCE, KENNETH W. "The Nature of Theory Construction in Contemporary Psychology," *Psychological Review,* 51 (January 1944) : 47–68.

———. "Relation of Learning Theory to the Technology of Education," *Harvard Educational Review,* 29 (Spring 1959) : 84–95.

STOGDILL, RALPH M. *Individual Behavior and Group Achievement: A Theory, the Experimental Evidence.* New York: Oxford University Press, 1959.

SULLIVAN, JOHN W. N. *The Limitations of Science.* New York: Mentor Books, 1949.

TATE, MERLE W. "Operationalism, Research, and a Science of Education," *Harvard Educational Review,* 20 (January 1950) : 11–27.

THORNE, FREDERICK C. "Critique of Recent Development in Personality Counseling Theory," *Journal of Clinical Psychology,* 13 (July 1957) : 234–44.

VAN DALEN, DEOBOLD B. "Relationship of Fact and Theory in Research," *Educational Administration and Supervision,* 45 (September 1959) : 271–4.

WHITE, ROBERT W. "Motivation Reconsidered: The Concept of Competence," *Psychological Review,* 66 (September 1959) : 297–333.

In a population which is so dependent upon research, it is sad to reflect how few people perceive what it is all about.

PALMER O. JOHNSON

4 The Steps of The Scientific Method

As Americans we are justifiably proud of our many scientific advances. Not only has science helped us attain a position of international prestige and supremacy, but it also has provided us with the highest standard of living in the world. Our scientists are among with the world's best in their fields of specialization, and our factories employ the latest technological devices to produce a gross national product unequalled anywhere. At the materialistic level, we are truly a scientific nation.

Unfortunately, our claim to science as a personal attribute, an attitude, and a way of life is not so indisputable. The average American still harbors misconceptions, superstitions, prejudices, and numerous other unscientific notions. Too frequently, he reveals that his scientific attitude is really quite superficial. Not only is he governed more directly by hunches, feelings, and opinions than by facts, but he is not sure of what science is—either as a product or as a procedure.

Even our high-school and undergraduate students, despite two or more years of "science," generally have only a superficial (if not erroneous) conception of its nature. They frequently equate science with dissecting frogs or with "discovering" the chemical composition of compounds. Rarely are they suffi-

ciently aware of the unity of science or of the philosophical and sociological setting in which it operates.

Although the objectives of science education have been formulated by a number of writers,[1,2] they are, unfortunately, inadequately incorporated into science courses. Too often students are introduced to the glamor of science by a lab manual with step-by-step directions for reaching the predetermined answer to their "problem." The textbook is too frequently the same unconditional authority in science classes as it is in other classes—with its contents equally, if not more, indisputable. There is a definite need for science teaching that ensures a greater understanding of the tentative and relative nature of scientific laws, of the need for flexibility in the application of the scientific method, and of the nature of science as something to be lived rather than something to be learned. It is sometimes disconcerting to see students who use opinion and fact with the same tone of dogmatic finality, who cannot tell one from the other, and who cannot substantiate an opinion except by locating another individual who holds the same views, and who equate consensus with truth.

Although graduate students in education are generally well up on their "facts" few of them appreciate the scientific basis on which education must rest. Research is often considered a competitor of, or a substitute for, constituted authority rather than a means of discovering knowledge. Too frequently research is viewed as a formalized process of applying a rigid sequence of steps to the solution of a problem. If we are to gain maximum benefit from our present orientation of education toward science, teachers must come to see that science is simply a matter of disciplined common sense, and, more important, that the answers to educational problems must come from systematic research.

THE RESEARCH PROBLEM

Selecting a Topic

Probably no aspect of graduate study is more unnerving to the student than the selection of a research topic, whether the

[1] Mary A. Burmester, "Behavior Involved in the Critical Aspects of Scientific Thinking," *Science Education*, 36 (December, 1952) : 259–63.

[2] Leo Nedelsky, "Formulation of Objectives of Teaching in the Physical Sciences," *American Journal of Physics*, 17 (September 1949) : 345–54.

research is a formal requirement for the thesis or a project in a course in research. Unfortunately, the student is frequently expected to select a topic early in his graduate work, at a time when he is not ready for such a selection. Not only is he unfamiliar with the nature of research itself, but generally he is also unsure of the areas in which research is needed and of the procedures he is to follow in getting relevant answers. To make matters worse, he finds his range of selection restricted by his lack of competence in the more advanced statistical techniques necessary for dealing adequately with the more significant educational problems.

Too frequently, after anxious conferences with his advisor, the student "chooses" a topic suggested by the latter because any topic is better than no topic at all. Some students, after many hours of exploration, abandon their topic and start afresh, while others continue despite the unsuitability of their problem and end up having nothing worthwhile but the satisfaction of having met another requirement. As a result, the thesis, which should be the most rewarding experience of graduate work, becomes sheer drudgery, and the degree itself becomes the goal of graduate work.

Inability to select a topic is a common weakness of graduate students. Frequently even students who show exceptional competence in classwork somehow lack whatever is involved in the undertaking of a major project on their own. Unfortunately, the present graduate education program is so organized that the student's first attempt at an individual research project comes at the very end of his program, when it seems a little late to be recognizing weaknesses.

Although it is apparent to all college advisors that the field of education is "just bristling with problems" to be selected and solved, no such clarity of insight is given the poor student, who too often finds that all of his ideas about research topics fall in the category of "too big," "too small," "already done," "incapable of solution," "beyond his resources and talents," and so on. The little anecdote by Buckingham[3] reproduced in part below describing the case of the student who comes to "discuss" a thesis problem, and who finally retires "a

[3] Editor, "The Editor Turns Professor," *Educational Research Bulletin*, 6 (September 14, 1927) : 252–3.

discouraged seeker after truth in a world where all the problems have been solved" is, unfortunately, too commonplace

"I've got to write a Master's thesis," says he, "and I'd like to talk to you about a topic." The statement ends with a slight upward inflection as if, in spite of its grammatical form, a sort of question were implied. After an awkward pause Mr. Blank (the student) repeats that he would like to talk about a thesis topic. Whereupon the editor (and professor) suggests that he go ahead and do so.

It transpires, however, that the editor-professor has misconceived Mr. Blank's meaning. He has no topic to talk about. In fact, instead of coming with a topic, he has come to get one. He looks so expectant, too; purely, as one might say, in a receptive mood.

No, he has no problems to suggest. He gives one the impression of having just learned about this thesis business, and of being entirely open-minded on the subject. At least, one gathers that he has no bias toward any particular topic and certainly no preconceived notions. (Graduate professors will recognize this as a familiar situation.)

A conversation ensues. The editor—playing for the nonce his professorial role—asks in what department Mr. Blank is majoring, what courses he has taken, what positions he has held, and for what type of educational service he is fitting himself. At one stage of the resulting exchange of ideas Mr. Blank brightens. With some modesty, yet with the undeniable air of a discoverer, he suggests that he might correlate intelligence and achievement in the high school. He could give some tests in the school with which he is connected; and his friend, the principal of the X school, would probably let him give some tests there; and maybe he could get one or two more schools if he stopped to think about the matter. And, O yes! how many schools does the professor think would be needed to get results that you could depend on? On being told that intelligence and achievement—so far as either is now measurable—have already been correlated by hundreds of people, Mr. Blank helplessly withdraws within himself, a discouraged seeker after truth in a world where all the problems have been solved.

Actually, the plight of the student described above is not one of complete fabrication. The adequate problems with which education is faced are frequently of such magnitude that they would not be suitable topics even for a doctoral dis-

sertation, and though the anecdote may be both humorous and depressing, it can perhaps be appreciated when the overall situation is taken into consideration. On the other hand, this is an area of primary importance, for the secret of success in research is frequently as much a matter of selecting appropriate problems as it is of being able to solve the problems that have been selected. Furthermore, this is an area in which a student is really on his own, and this is precisely the purpose of the thesis or dissertation requirement, for it is here that the aspirant to high professional status shows if he is capable of demonstrating the necessary initiative, originality, and good judgment. It should be his prerogative, rather than simply his responsibility, to select his topic, to plan its investigation, and to derive its solution, drawing on outside help only in emergencies and for confirmation of the decisions he has made. He can, of course, draw on the experience of his advisor and his major professors, but he must not lean on them for carrying out his study. Even when he is working on a problem that is part of an overall research program, the graduate student must limit his requests for advice to what is necessary to co-ordinating his efforts with the overall project, rather than cast himself in the role of a clerical assistant.

Duplication

Unfortunate indeed is the graduate student who finds that the problem in which he has invested time and effort has already been solved or is the object of a prior claim, for a basic consideration in the choice of a research topic is the avoidance of duplication. It therefore behooves the prospective investigator to survey the literature carefully before he begins his study to ensure that his problem has not already been solved to the point that his contribution over and above that which has already been discovered would be relatively trivial, or that it is not already under investigation.

The interpretation of what constitutes duplication is, however, a matter of some debate. It is recognized that some studies can be repeated profitably either to check their validity or to extend the applicability of their conclusions. Duplication is acceptable, for instance, if the student can bring new evidence to bear on the subject by using an improved design, or if

changed circumstances make verification under the new conditions desirable. For example, the four-quarter school year, the subject of numerous studies in the twenties and thirties, would bear re-investigation now that air-conditioning has minimized the debilitating effect of summer heat and that increased emphasis on family travel has complicated the co-ordination of summer vacations. In fact, many of the problems solved many years ago are in need of verification, extension, or re-evaluation now that newer and better psychometric tools, research designs, and statistical procedures are available.

The question of duplication must be considered in the light of the principle that, to be acceptable, a thesis or dissertation must make a contribution to the advancement of education as a science. There are educators who feel that the master's student will never make any great contribution to anything, and that he might as well be allowed to repeat a good study merely as an exercise in scholarship. They point out that the taboo on duplication deters students from investigating any problem which has been studied before, thus depriving them of many opportunities for significant investigations and resulting in a lack of continuity in the research on a given problem. It is their feeling that master's candidates probably are best employed in respading the grounds from a number of angles. These contentions have some merit in suggesting a somewhat liberal interpretation of what constitutes duplication, but they probably do not justify mere repetition. Although thousands of studies are conducted every year, problems are not getting any scarcer. On the contrary, every study that provides a tentative answer to a problem simply uncovers a multitude of other problems that need investigation. There is, therefore, no point in going over what is already known when there is so much new territory to be explored.

Of course, this does not deny the fact that rarely are problems in education solved with such finality that further verification is unwarranted. Furthermore, rarely are all the aspects of a given problem solved, and it sometimes is possible to carry the investigation of a problem area beyond the first study to the next step. Problems under investigation are also frequently fruitful sources of suggestions for parallel studies that could be conducted in related areas without involving objectionable duplication.

Criteria for Selection of a Research Topic

Although there are no standard rules that, either singly or collectively, will guarantee the suitability of a research problem, a number of criteria in the sense of necessary—though not sufficient—conditions might be listed for guidance in the selection of a topic.

1. Is the topic of interest? While interest sometimes develops with familiarity, it does not seem likely that the student can do his best work on a topic that has no personal meaning for him.
2. Is the topic sufficiently original that it does not involve objectionable duplication?
3. Is the topic amenable to research? Many problems are of a philosophical nature; they can be discussed but not to the point where objective evidence can provide a solution. Thus, the problem, "Should high school boys work?" having no referent is, as stated, a philosophical issue not subject to scientific determination. Before it could be investigated, it would have to be oriented toward a criterion—for example, "Do high school boys who work suffer academically?"
4. Is the problem significant? Specifically, what will it add to the present state of knowledge or the development of education as a science? There are so many problems that need to be investigated that it does not make sense to have a student devote himself to the study of a trivial or nonsensical problem. Wolfle, for instance, ridicules the trivial topics that have been subjected to research. Referring to Longfellow's poem, "I shot an arrow into the air; it fell to earth I know not where. I breathed a song into the air; it fell to earth I know not where," he points out that some people would not go about the task so light-heartedly; they would want to have a control group of poets who do not breathe songs into the air—or would want to conduct a survey to determine the differential effects of song-breathing and nonsong-breathing poets, for example.[4]
5. Is research into the problem feasible? Are data available in the situation in which the investigator finds himself? A very significant contribution to science could certainly be made by investigating the existence of life on Mars, for example, but such a project is not feasible at this time. Many significant educational problems also must be by-passed, perhaps because they would not get the necessary clearance or because they

[4] Dael L. Wolfle, "Fetish of Experiment," *Science*, 125 (1957): 177.

are not possible from the standpoint of the competence of the investigator.

Sometimes, despite all precautions, the problem selected turns out to be unsuitable. No matter how annoying this may be, it is generally better for the investigator to abandon his project and to move on to a different problem. Although an unsuitable problem sometimes can be converted into one of a similar or parallel nature, and part of the effort salvaged, it is usually foolish to continue to invest time and effort in a project discovered to be marginal.

Sources of Problems

Although there are a number of sources from which leads to the selection of a problem can be obtained, there is no standard prescription that can be given that will provide every student with a suitable problem. Nor is there a standard source from which a student can simply choose a topic. There is no alternative but to become a scholar in the field, to know what the problem areas are, and to use imagination.

Research problems can sometimes be located from reading the professional literature. If the student reads critically, he can find points on which he disagrees with the author, or he can locate studies with results that can be challenged or, at least, that need to be verified, or some studies conducted under a one set of conditions rather than under other equally legitimate circumstances. As he pursues his studies, a student is bound to find many gaps in the contents of a given subject pointed out by the author of his textbooks or by the instructor of his classes. Suggestions for problems are frequently found in the articles of the *Review of Educational Research* or in the *Encyclopedia of Educational Research,* as well as in other journals. Reading the professional literature often suggests the possibility of conducting parallel studies in different fields or with different populations. It also may be possible to combine two ideas into a single study.

Once the student has located the area in which he wants to work he may find it helpful to peruse the articles listed in the *Educational Index* for suggestions. Yearbooks of the various societies are also particularly helpful in listing some of the problem areas in which research is needed. Not to be over-

looked in the search for a topic is the student's personal experience and the situation in which he finds himself. In the average school system an alert teacher can find many areas which need investigation. Also helpful are the suggestions of administrators, supervisors, and other persons of wide background. Administrators often lack the means and personnel necessary for research into some of their problems and might welcome the research efforts of the graduate student.

It has been suggested that educational research must not be restricted to the solution of immediate problems at the empirical level, but rather must also be oriented toward the integration of research findings into a conceptual framework that gives them meaning and broad usefulness. There is nothing unscientific, however, about solving a practical problem, and it is sometimes better for the graduate student to work on such a problem than to attempt to deal with the more complex organizational or theoretical aspects of science, which frequently require a greater insight into the overall field than he is likely to possess.

Choosing the Topic

Some people are more sensitive than others to the existence of problems and are, therefore, more capable of selecting appropriate research topics. Probably the two major factors involved in selection are experience and creativity. It seems logical to expect that good problems stem from a clear understanding of the theoretical, empirical, and practical aspects of the subject, derived from personal experience and from a thorough review of the literature. Conversely, lack of familiarity with the subject is almost sure to result in a poor choice. For example, a little sophistication in the area of intelligence and its measurement would probably restrain the student from attempting to isolate the relative effects of heredity and environment on intelligence.

The second major contributor to the wise choice of a problem is creativity and the other personality factors that make for originality, flexibility, initiative, ingenuity, and foresight. These attributes must operate within the framework of what is already known, and, generally, familiarity with a given field is conducive to original thinking. The contrasting view-

point is that too thorough an immersion in the literature of a given field is likely to blunt originality and force the individual's approach into a standard mold. This, of course, need not be so, particularly if a person reads with a view toward critical analysis of what is read rather than toward simple acceptance and absorption. Of course, the student must also recognize that, though faculty advisors are invariably delighted when a student takes the initiative and locates a suitable problem on his own, the acceptability of a topic is a decision for the advisor and his committee, and there are bound to be occasional differences of opinion as to what constitutes an adequate research problem. While this arrangement exerts an influence towards conformity which is perhaps stifling in rare instances, it usually works to the student's benefit.

It may be of advantage to the student seeking a problem for investigation to attempt to structure the field on the basis of such questions as those presented by Holmes *et al:*

1. In your field of interest what practical problems have to be met by those individuals who do the actual work?
2. In current and recent research, what problems are under active attack?
3. What facts, principles, generalizations, and other findings have resulted from research in your field?
4. What practical implications for schoolwork may be drawn from the results?
5. To what extent have the findings of research actually been applied in your field?
6. What problems remain to be subjected to research and what problems are now emerging?
7. What are the chief difficulties to be met in prosecuting the researches yet to be conducted in your field?
8. What are the interrelations between research in your field and research in adjacent fields?
9. What research techniques or procedures have been developed in your field?
10. What concepts have been operative, either explicitly or implicitly, in the research in your field?
11. What assumptions have been implicit or openly avowed in the research in your field?[5]

[5] Henry W. Holmes, *et al., Educational Research: Its Nature, Essential Conditions, and Controlling Concepts.* (Washington, D.C.: American Council on Education, 1939), p. 51ff.

Clarifying and Stating the Problem

Before the student can attack his problem effectively, it is essential that he clarify its nature. A vague problem is more likely to lead to untold difficulties than to significant outcomes. A common error of the beginning student is choosing a topic that is too broad. For example, the problem of "teacher effectiveness," taken in the broad sense, is more than one master's or even doctoral student can investigate effectively. Such a study needs to be delimited to selected aspects of the total picture that can be isolated meaningfully. Conversely, the problem selected must not be so narrow that it becomes artificial—for example, a study of the effects of writing posture upon penmanship would be meaningless. In practice, it is better to begin with a semibroad problem, and, as one proceeds to review the literature, gradually to restrict it. The major delimitation should, of course, take place before the data are collected and, to some extent, even before the literature is surveyed in detail.

The variety of errors that can be made in the formulation of research projects is relatively unlimited. A common fault, for example, is to list a field or broad area rather than to state a specific problem. A study of "Juvenile Delinquency" or of "Teacher Effectiveness" would be more feasible, if it were restricted to a comparison of the personality of delinquent and non-delinquent boys, or of the professional attitudes of "good" and "poor" teachers. Another common fault is to state a problem in such a way that its investigation is essentially impossible —such as "The desirability of introducing typing in the elementary school," or "The effects of working mothers on the academic achievement of their offspring."

Following are a few examples of problems that have been reformulated into somewhat more feasible projects. They perhaps could be altered further and are presented simply as illustrations.

PROBLEM: The role of the principal in American public education.

RESTATEMENT: The supervisory practices of principals in large high schools of . . . City.

PROBLEM: A survey of factors affecting pupil progress.

RESTATEMENT: A survey of the level of aspiration of over-achievers.

PBOBLEM: The use of tests in college admission.

RESTATEMENT: The prediction of academic success at College X.

PROBLEM: The relation of socio-economic status to intelligence.

RESTATEMENT: A comparative study of the performance of children different socio-economic status on the Stanford-Binet.

PROBLEM: A study of the effectiveness of remedial reading courses at the college level.

RESTATEMENT: A comparative study of three methods of improving reading speed and comprehension among college freshmen.

PROBLEM: The value of a remedial reading problem at the college level.

RESTATEMENT: A study of the effects of a remedial reading program on the academic achievement of college freshmen.

PROBLEM: A study of the factors that relate to college achievement.

RESTATEMENT: A study of the effect of part-time employment on the scholastic achievement of freshmen women at College X.

A good grasp of the problem should provide the student with insight into what can be done in the research study he is contemplating, not only in defining his problem but also in deriving hypotheses and likely methods of attack. Such a background can generally be obtained from a thorough review of the literature. If the operation of the variables involved is not known, however, it is highly desirable to conduct a pilot study in order to clarify both their nature and the means of their investigation before the final statement of the problem is made. This, of course, takes time, but it is invariably a wise investment, in that it provides greater insight into the nature of the problem and permits its more precise and adequate formulation.

If the problem is to serve as a guide in planning the study and interpreting its results, it is essential that it be stated in precise terms. Only then can it give direction to the collection of the data and to the manner in which they must be processed in in order to provide the required answer. The student often

is impatient to get started and is likely to forget that a study generally is only as good as the clarity with which the problem has been stated. Without such clarity, he is not likely to know what data he is to collect, or how to relate the data which he has collected to his problem. Particularly to be avoided, for example, are meaningless clichés and verbalisms which have no relationship to measurable operations—such as "the effects of emotional security upon the child's all-round growth."

The procedures to be used in solving the problem also must be defined clearly. Such a definition must be made when the problem is selected inasmuch as it may be found that there is no way of solving the problem as stated, and that it will have to be restated according to what is feasible from the standpoint of method. Thus, sometimes it is the nature of the data which can be collected that determines the problem that can be selected for research. It is almost impossible, for example, to compare the relative effectiveness of the newer School Mathematics Study Group program with the traditional approach to the teaching of mathematics, simply because they do not cover the same content and are not oriented toward the same immediate objectives.

In the final analysis it is the problem as defined that determines the data that need to be collected, and only data that fit the framework of the problem as stated should be collected. It follows that the whole problem must be explicitly defined—from the standpoint of both the specific question to be answered and the techniques to be followed in providing the required answer—before any attempt is made to gather the data.

There is no standard form for the presentation of a problem. Some schools insist that it should be stated in the form of a question, others in the form of a hypothesis to be tested, and still others in the form of a statement. The form is relatively inconsequential. What is important is that the problem be stated in such a way that both the investigator and the reader know precisely what is to be investigated, and how this is to be accomplished.

THE HYPOTHESIS

The Nature and Purpose of Hypotheses

The derivation of a suitable hypothesis goes hand in hand with the selection of a research problem. A hypothesis can be

considered a tentative generalization about the problem under investigation. It is an assumption or proposition whose tenability is to be tested on the basis of the compatibility of its implications with empirical evidence and with previous knowledge.

Modern investigators are agreed that, whenever possible, research should proceed from a hypothesis, for, in the words of Van Dalen, "a hypothesis serves as a powerful beacon that lights the way for the research worker."[6] Hypotheses are particularly necessary in studies where cause-and-effect relationships are to be discovered. They are perhaps less crucial in studies in which the task is one of determining the status of a given phenomenon, although even in such studies the investigator is likely to need some tentative hypothesis to guide him to the areas worth exploring. Actually, hypotheses are not essential to research, particularly in the early stages of the exploration of a problem. Scientific discoveries can emerge from investigations not directed by hypotheses, and, though hypotheses are generally useful guides to effective research, it must not be assumed that failure to have a hypothesis is necessarily a sign of a lack of scientific orientation.

The objection to beginning with a hypothesis is that postulated by Bacon, who felt that a hypothesis biased the investigator toward a given position and caused him to lose his objectivity. This need not be so: a hypothesis must be conceived as an assumption which merits consideration, not as a position to be defended. Furthermore, the scientific method puts such restrictions on the investigator that the extent to which he can distort the evidence to fit his personal views is minimal. While it may be true that hypotheses can blind the investigator to other more fruitful hypotheses and cause him to ignore data which are not compatible with his hypothesis, this is the exception rather than the rule in good research.

Actually, it is almost impossible for a person who has a clear picture of his problem not to have one or more hypotheses more or less clearly in mind. The only question, therefore, is the degree to which those hypotheses are recognized at the conscious level and elaborated, screened, refined, and, finally, used

[6] Deobold B. Van Dalen, "Role of Hypotheses in Educational Research," *Educational Administration and Supervision*, 42 (December, 1956): 457–62.

as a pivot around which the investigation is to center. Conversely, if the investigator is not capable of formulating a hypothesis about his problem, he may not be ready to undertake its investigation. As Burton points out,[7] the derivation of the hypothesis should precede the collection of the data. This is indisputable in the usual case, but only to the extent that the investigator of a given topic would generally have enough background for him to derive intelligent, albeit tentative, hypotheses.

The arguments in favor of well-developed hypotheses as a framework for research center around the fact that the aimless collection of data is not likely to lead anywhere. Since a multitude of possible relationships can exist among phenomena, generalizations and relationships significant from the standpoint of a given problem do not just emerge from data. More specifically:

1. Hypotheses provide direction to research and prevent the review of irrelevant literature and the collection of useless or excess data. Hypotheses define what is relevant and what is irrelevant, since facts derive meaning only when considered in the light of meaningful hypotheses. They enable the investigator to classify the information he has collected from the standpoint of both relevance and organization, for a given fact may be relevant with respect to one hypothesis and irrelevant with respect to a second, or it might belong to one classification with respect to the first hypothesis and to an entirely different classification with regard to the second. Hypotheses not only prevent waste in the collection of data, but also ensure the collection of the data necessary to answer the question posed in the statement of the problem.
2. Hypotheses sensitize the investigator to certain aspects of the situation which are relevant from the standpoint of the problem at hand. In general, hypotheses spell the difference between precision and haphazardness, between fruitful and fruitless research.
3. Hypotheses are not ends in themselves, but rather are the means by which the investigator can understand with greater clarity his problem and its ramifications, as well as the data which bear on it. They enable a researcher to clarify the pro-

[7] William H. Burton, *et al.*, *Education for Effective Thinking.* (New York: Appleton-Century-Crofts), p. 62.

cedures and methods to be used in solving his problem and to rule out methods which are incapable of providing the data necessary to test the hypothesis posited.

4. Hypotheses act as a framework for the conclusions. They permit the collection of relevant data, and they make possible the interpretation of these data in the light of the potential solution. Hypotheses provide the framework for stating conclusions in a meaningful way—that is, as a direct answer to the hypothesis being tested. In fact, it may be better when dealing with complex phenomena—such as teacher effectiveness—to have multiple hypotheses, each suggesting its own criterion and its own means of solution in order to encompass the problem on all sides, and thus ensure its more complete appraisal and resolution.

Sources of Hypotheses

The task of deriving adequate hypotheses is essentially parallel to that of selecting suitable problems, since the selection of a problem can hardly be considered apart from the hypothesis that might be tested in its solution. And just as there is no royal road to the location of a suitable problem, there is no royal road to the discovery of fruitful hypotheses. There is also a parallel in the characteristics of experience and creativity that make certain persons capable of deriving adequate hypotheses. And, though hypotheses should precede the gathering of data, a good hypothesis can come only from experience. Some degree of data-gathering, such as the recall of past experience, the review of the literature, or a pilot study, must therefore precede the development and gradual refinement of the hypothesis. It would be difficult, for example, to derive meaningful hypotheses regarding the various aspects of teacher effectiveness without some background in the psychology of learning as well as in the effects of such teacher characteristics as sex, age, experience, and training on pupil growth.

The factor of persistence must not be overlooked. Success at discovery is invariably predicated on the expenditure of considerable time and effort in tracing various leads and refining tentative hypotheses. Unfortunately, the general pattern is for the investigator to report his final hypothesis and the success to which it led; he never mentions the dozens of hypotheses which he discarded. As a result, other investigators may waste

time on the same fruitless leads, though to be sure, hypotheses which have been discarded too quickly may prove useful when approached from a different point of view.

Actually a good investigator must have not only an alert mind capable of deriving relevant hypotheses, but also a critical mind capable of rejecting faulty hypotheses. Interestingly enough, the person who is "full of ideas" may also be the person who is lacking in critical analysis—that is, originality may be somewhat incompatible with a critical attitude.

Although reasoning by analogy generally is considered unacceptable as a source of proof, it is a very fertile source of hypotheses. The premise is that, if the two situations are alike in certain aspects relevant from the standpoint of the problem under consideration, they are probably similar in other relevant aspects. It is assumed that the existence of similarities between two situations is not accidental, but that it is the result of the operation of some law common to the two situations so that the other similarities governed by the same law obtain in both instances. Analogy is never based on complete likeness, but the differences are assumed to be in those aspects which are independent of the common law and which therefore can be ignored. This, of course, cannot be shown through logic, and reasoning by analogy is suspect unless and until its outcomes have been verified through empirical proof. Nonetheless, the insights that analogy provides are useful inasmuch as they lead to their own refinement and verification through the acquisition of relevant data and the formulation of more adequate hypotheses.

Criteria for Judging Hypotheses

The relative merits of a given hypothesis can be judged only by its effectiveness in the particular problem under investigation, and its final validity cannot be appraised except through an empirical test. Nevertheless, one can set up certain general criteria on the basis of which to judge the relative worth of a hypothesis. (These criteria, it will be noted, parallel rather closely the criteria of a good theory presented in Chapter 3.)

1. A good hypothesis must be based directly on existing data. It might even be expected to predict or anticipate previously unknown data.

2. A good hypothesis must explain existing data in simpler terms than any competing hypothesis. The law of parsimony favors the hypothesis that explains the most in the simplest terms.
3. A good hypothesis must be stated as simply and concisely as the complexity of the concepts involved will allow. Note, for example, how simply some of the major laws of science—gravity, motion, survival of the fittest, and others—are stated.
4. A good hypothesis must, above all, be testable. It must be stated so that its implications can be deduced in the form of empirical or operational referrents with respect to which the relationship can either be validated or refuted. For instance, if the world is round, one can reach the East by sailing West. The statement of the hypothesis must permit the development of a research design capable of providing the data necessary for testing its validity. For example, the basic premise of the Montessori system of education—that freedom of movement within the classroom is an essential condition for effective learning—lent itself readily to an empirical test. On the contrary, the hypothesis that a permissive environment is conducive to the all-round growth of the child is relatively untestable because of the lack of precision with which "permissive environment" and "all-round growth" can be defined and measured. The proposition that kindergarten promotes social and emotional maturity is likewise difficult to test because of the relative unavailability of adequate means for the valid appraisal of such maturity and for the isolation of other factors which also contribute to social and emotional growth. It also must be recognized that pointing to a correlation between the variables in question generally does not constitute an adequate test of a hypothesis.

Testing the Hypothesis

The proof of the worth of a hypothesis lies in its ability to meet the test of its validity. Validity is established in two stages: 1. The statement of the hypothesis allows the investigator to develop deductively certain implications which, when stated in operational terms, can lead to the rejection of hypotheses that are in conflict with accepted knowledge at the logical level. For example, the hypothesis that marbles of different weight rolling down an incline onto a horizontal platform would roll distances proportionate to their weight would have

to be rejected, since it conflicts with Galileo's findings that the rate of fall is independent of the weight of the object. 2. If a hypothesis passes the test of logic, it then must be subjected to an empirical test, perhaps through an experiment or a series of measurements. The hypothesis that boys are stronger or taller than girls, for example, can be verified through measurements. In a complex study, such as the comparison of the relative effectiveness of different combinations of instructional procedures and class size, a research worker may consider three or four hypotheses simultaneously, each to some degree different from the others.

A hypothesis is never proved: it is merely sustained or rejected. If it fails to meet the test of its validity, it must be modified or rejected. A hypothesis can be useful even if it is partly incorrect, however, and, in practice, hypotheses are not so much rejected as they are replaced by more adequate hypotheses. Usually the negative instances which occur require only further clarification and refinement of the hypothesis rather than its outright abandonment. Thorndike's hypothesis concerning the role of practice in promoting learning, for instance, was later integrated in the present version of the *law of effect*. Negative instances suggest the presence of other considerations which must be isolated or incorporated in the statement of the hypothesis so that the exceptions can become part of the relationship at a more sophisticated level.

The confirmation of a hypothesis, on the other hand, is always tentative and relative, subject to later revision—and even rejection—as further evidence appears or as more adequate hypotheses are introduced. Logical and empirical verification can never provide conclusive proof, and confirmation must always be a matter of probability rather than of certainty. This is essentially the pattern which we noted in connection with theories, though, to be sure, hypotheses, since they are more tentative and less fully developed than theories, are more subject to modification and to rejection.

Hypotheses, Laws, and Principles

When a hypothesis is sustained by logical and empirical tests, it provides the basis for generalizations or conclusions. As further confirmation and clarification of the conditions under

which the hypothesis holds accumulate, a generalization, if its importance warrants it, may become a law or principle. The distinction between a hypothesis, a generalization, a law, and a principle is generally a matter of *dependability,* based on such factors as logical and theoretical plausibility, repeated verification, and adequate definition and delineation of the conditions under which it holds; and *complexity, scope* and *relative importance.* Thus, Galileo probably began with a simple hunch (*hypothesis*) that the rate of free fall of a body is independent of its size and weight. A few confirming instances may have led him to a generalization (*conclusion*). Later as its importance and scope became recognized, the discovery was given the status of *principle* or *law.* The point at which the transition from one to the other takes place is, of course, imprecise. The terms *law* and *principle* are generally used interchangeably to refer to the statement of an invariant—as far as it is known at the present—relationship among phenomena. The concept of parsimony, for example, is variously referred to as a law or a principle. Technically, however, a principle is more comprehensive than a law and may serve as a basis from which laws are derived. In mathematics, for instance, the term *principle* is frequently used as a synonym for *axiom.*

A scientific law may be defined as a hypothesis whose scientific validity is relatively unquestioned. It represents as close an approximation to empirical truth as has been derived to date, although, to be sure, every year a number of "laws" have to be recalled for revision and extension, and perhaps, rejection. In fact, it may be suspected that most laws, at least in their original statement, incorporate some degree of error and/or incompleteness. As more and more data are accumulated, laws become progressively broader in application, covering more and more of the known aspects of phenomena, more and more adequately. In their later stages, laws are best explained as the logical outgrowths of theories. Thus science involves a progressively more adequate explanation of events and phenomena by a complex of more and more adequate hypotheses, laws, principles, and theories, logically interrelated into a meaningful whole.

As we noted in Chapter 3, laws may be *nomothetic*—refer-

ring to relationships common to all individuals of a given set—or *idiographic*—pertaining to the individual case. Laws can be classified further as *empirical*—for example, water boils at 212° F—and *theoretical*—for example, PVT is a constant.

Although the distinction between theories and hypotheses is not always clear or free of overlapping, a theory is broader in scope and rests on a somewhat more sophisticated basis than does a hypothesis. Thus, while a hypothesis may be postulated on the basis of a relatively haphazard observation or a relatively unimportant phenomenon, a theory generally attempts to unify a number of previously established generalizations. This is, of course, most evident in such advanced theories as the theory of evolution or the theory of relativity.

THE COLLECTION AND ANALYSIS OF DATA

Scientific problems can be resolved only on the basis of data, and a major responsibility of the scientist is to set up a research design capable of providing the data necessary to the solution of his problem. While the unity of research makes it impossible to say that one aspect is more crucial than another, the collection of data is of paramount importance in the conduct of research, since, obviously, no solution can be more adequate that the data on which it is based.

The more clearly and thoroughly a problem and its many ramifications are identified, the more adequately the study can be planned and carried to successful completion. Thus the task is to synchronize the statement of the problem with the design to be used in its solution, and every aspect of the study down to the last detail of execution must be planned before the study is undertaken. It is senseless to select a topic, no matter how adequate, if circumstances preclude the collection of the data required for its solution. And, of course, the student who leaves the statistical treatment of his data for "When I get there" may find that the data, as collected, are impossible to analyze

The problems involved in the accumulation of adequate data are far too numerous and too technical to be discussed here; the discussion, therefore, will be restricted to a brief overview of the fundamental aspects of measurements as they relate

to research. More adequate treatment can be found in texts in educational and psychological tests and measurements, and the student is referred to such sources with the reminder that the field is of primary importance. No one interested in research can afford to be without thorough training in this area.

The Nature of Data

Data can be classified into two broad categories: *qualitative data* or *attributes*—for example, color, intelligence, and honesty—and *quantitative data* or *variables*—for example, IQ, grade-point average, and height. The distinction frequently is based on processes rather than on properties inherent in the phenomena, for generally properties considered qualitative can be made quantitative by measuring them with an instrument designed to assign numerical values to the various degrees to which they exist. Thus, intelligence, height, personality adjustment, and so on exist both as attributes and as variables. As a result, the decision to research a given phenomenon on the basis of its attributes, or on the basis of its quantitative aspects, is frequently a matter of choice, depending on such considerations as the need for precision and the ease of manipulation of the data. In general, the latter alternative is the more functional and the more adequate, since quantification provides a greater refinement in classification and possesses definite advantages over qualitative listings by virtue of its amenability to more adequate treatment by the modern statistical processes. In fact, the quantification of phenomena generally is considered essential to the progress of a science, particularly at the more advanced levels.

Unfortunately, at present we do not have the instruments necessary for the precise quantification of many of the characteristics with which educational research is concerned—for example, honesty, health, adjustment, or motivation. Although we are devising progressively more adequate techniques and instruments with which to "measure" what years ago existed only as attributes, we still have a long way to go, particularly in the more intangible aspects of human behavior. It appears that by their very nature certain properties and characteristics—for example, such concepts as married, widowed, and dead,

—must remain attributes for which the only quantification possible is that of counting the frequency of occurrence. On the other hand, it may be possible to convert even such attributes as sex into what is probably the more meaningful psychological dimension of "masculinity-feminity," and thus convert what is essentially a dichotomous attribute into a measurable quantity.

Variables

Variables can be classified as *continuous* or *discrete*. Continuous variables are those for which fractional values exist and have meaning—for example, distance, age and weight, where 4.8712 miles, 68.117 years, or any other fraction of a whole unit is logical and measurable within the precision of the instrument used. Discrete variables, on the other hand, exist only in units (usually units of one). There are 29, 30, 31, . . . students in a class, 800, 801, 802, . . . volumes in a library, and so on. Here, fractional values cannot exist; one cannot have 11.25 eggs in a basket; nor can a couple have 3.25 children, in the usual sense of the words *eggs* or *children*. This distinction is somewhat more complicated in practice: What should a college, reporting its enrollment, do with students carrying a half-load? How does a library enumerate three booklets bound into one volume, since it could have as easily had three volumes by binding each separately? The problem is generally resolved—though not entirely satisfactorily—by defining the unit of operation. Thus, the library would have to indicate whether it is referring to the number of volumes separately indexed in the card catalogue or to the number of separate titles as listed in the *Cumulative Book Index*, or it might possibly present the data both ways.

The typical problem in educational research deals with test scores. These are generally reported as discrete variables, though they are often fundamentally continuous. Thus, a child having 19 out of 20 words correct on a spelling test gets a score of 19, but inasmuch as he may have missed the twentieth word by a "country mile" or by a mere slip, it is possible to conceive of his true score as ranging anywhere from 19.00 to 19.99. IQ's also are recorded as discrete, though, by their very computa-

tion, they are technically continuous. It should be noted in this connection that a major reason for avoiding fractional values in many instances is that the accuracy of measurement does not warrant consideration of fractional values, not that the variables are constitutionally discrete.

In research, where the concern is with group values which almost invariably are fractional, continuous variables appear somewhat more acceptable than discrete variables. Thus, it seems to be more logically acceptable to think of the average distance traveled by commuters to City X as 9.28 miles, or of the average age of freshmen at College Y as 18.34 years, than to think of the average family as having 2.6 children.

Measurement

Man's first attempt to appraise the properties and characteristics of phenomena probably was made on the basis of a dichotomy. For instance, early attempts at studying the weather were probably restricted to noting whether or not it rained in a given day. Later this was probably extended to counting the number of times it rained in a given period, thereby providing a discrete series. The next step in the development of science was measurement, which provides a relatively unlimited number of categories into which phenomena can be ordered and which permits a more adequate and facile manipulation of the categories by virtue of their susceptibility to mathematical treatment.

Success in research, and in science, depends on the availability of instruments of sufficient precision to measure the phenomenon under study. Much greater progress in this connection has, of course, been made in the physical than in the social sciences. Most of the measurements with which educational research is concerned are derived through pencil-and-paper tests which are, as yet, relatively imprecise. This is particularly true in such areas as motivation, attitudes, values, and creativity, which are generally considered more psychologically and educationally significant than many of the variables which are being measured with greater accuracy. In fact, many educators would agree with Brown that the ease and accuracy with which educational outcomes are measured is frequently in direct proportion to their unimportance—what we measure most precisely

is precisely what it makes least difference whether we measure or not.[8] The problems connected with devising adequate instruments for measuring the more meaningful psychological dimensions and of processing the data which such instruments would yield are so complex, however, that their consideration here is inadvisable.

Characteristics of a Good Measuring Instrument

Measurement is effected by means of some instrument—for example, a gauge, a rule, a scale, or a test. If they are to provide dependable measurements, such instruments, regardless of their specific nature and purpose, must all possess certain qualities, of which *validity* is by far the most important—especially as the results apply to research. There is an important distinction to be made between measurement on an individual basis—such as in guidance—and measurement on a group basis. In the first situation, scores must be dependable individually; in the second, we are interested in group averages, and do not particularly object to individual errors—provided they cancel out.

A measuring instrument must be *reliable*—that is, it must be consistent in the measurement of whatever it measures. A test of intelligence, for example, would be lacking in reliability if in a test-retest situation, legitimately handled, a child's IQ shifted haphazardly from say, 70 to 140. Reliability is, of course, most important in guidance where the focus is on an individual child. In research, errors of unreliability, representing random errors, tend to cancel out so that in a fairly large sample, group values are not too greatly affected. Of course, a test relatively devoid of reliability—for example, an elastic yardstick to measure distance—cannot be used as the basis for scientific conclusions.

A good measuring instrument should also be *usable*. From an administrative point of view, usability is a consideration of practical importance. Research is especially concerned with usability because inadequacies in this area are readily transferred into errors of invalidity. Thus, if two groups are compared on the basis of their performances on a test of excessive length, the comparison will incorporate an element of motiva-

[8] Edwin J. Brown, "Some of the Less Measurable Outcomes of Education," *Educational and Psychological Measurement*, 2 (Winter, 1942) : 353–9.

tion and persistence which will confound the difference in the relative competence of the two groups. Similarly, excessive length in a questionnaire, for instance, is likely to result in a loss in validity, since it encourages non-response and thereby promotes non-representativeness in the returns.

Validity

Validity refers to the extent to which an instrument measures what it purports to measure. Operationally, an instrument is valid to the extent to which differences in test performance represent corresponding true differences among individuals in the characteristic the instrument is designed to measure. A test of history would be invalid, for instance, if it incorporated such a high level of reading proficiency that difficulties in understanding the vocabulary interfered with a student's performance on the test. A low score on such a test would not necessarily show a lack of knowledge of history, since the difficulty might have been in reading.

Failure to appreciate the importance ot validity is one of the most common errors vitiating research, particularly in the social sciences where validity is sometimes subtle and difficult to establish. It is frequently reported, for example, that a school is low in arithmetic competence, simply because its students did not perform at expected levels on a test bearing a title suggestive of arithmetic competence. It is essential to note that validity is a specific concept—a test is valid not in general, but is valid for a particular group under particular circumstances. The arithmetic test mentioned may not have been valid for the students of the particular school; tneir curriculum may not have been oriented toward the development of the competencies expected in the test. The norms of a standardized test are accumulated by administering the test to a large sample representative of the grade level or group for whom the test is intended. Since norms are simply standards of comparison, it is inevitable that a given class will not coincide with the norm group in every respect. In fact, the lack of equivalence of the two groups may be sufficient to account for the discrepancy in the performance of the class from that of the norm group.

Since validity is specific to a given situation, the legitimacy of the use that is made of a test cannot be considered apart from

the purpose for which it is being used. In an experiment where the purpose is to compare the relative performance of two groups, for instance, there may be no invalidity introduced by giving the two groups a short break even though this is in violation of standardization procedures. Such a step, however, would invalidate any comparison of the performance of the two groups with the norms of the test.

Test-wisedness on the part of children who have been tested repeatedly influences the validity of a test score in relation to the norms, and this factor is too frequently overlooked. Whenever the purpose is to compare the performance of a group against the test norms, it must be remembered that a test score is valid to the extent—and only to the extent—that the background of the testee is similar to the background of the group on which the test was standardized. For example, a high school, eager to have its graduates accepted into college, may encourage its students to take the College Boards two or three times during the course of their junior and senior years in order to get acquainted with the general nature of the tests and to orient their studies to the areas emphasized in the tests. Later the school may report that on the C.E.E.B. tests taken at the end of their senior year the students scored above national norms. To the extent that practice with the tests improved their performance, the scores made by students who have had greater than average contact with the tests would automatically be higher by an indeterminate amount than they should legitimately be.

A difficult problem which involves the concept of validity is that of the *fairness*—that is, the validity—of the instruments used to measure progress in an experiment. For example, the relative effectiveness of drill and the project approach in promoting academic growth may well hinge on the emphasis of the test on the basis of which this growth is measured. A comparison of the relative superiority of large *versus* small classes may also depend in no small measure on whether the criterion of the study is the memorization of facts or the development of critical thinking and favorable attitudes toward the subject. In such cases the investigator must determine what constitutes a valid criterion for the particular purpose of the investigation being conducted. Generally this requires a clarification of the objectives of the study, and a translation of these objectives

into a test or a series of tests representing a legitimate criterion of the comparison in question. Such an approach was particularly evident in the Eight-Year study. (See Chapter 15.) In any event, there is a need for a clear statement of the nature of the criterion with reference to which one method was found superior to another, since with reference to a different criterion, the relative superiority of the two methods might be reversed.

The concept of validity is not restricted to test scores; it applies to all data-gathering instruments and techniques. Thus, invalidity in research data might result from incompleteness of the returns or ambiguity in the items in a questionnaire study, the presence of the interviewer or the observer in an interview or observation study, the personal biases of the investigator, and so on. Sampling is another important consideration affecting the validity of the data gathered for research purposes (see Chapter 7).

THE INTERPRETATION OF THE DATA

The interpretation of research data cannot be considered in the abstract. In view of the diversity of the research methods used in education, and the corresponding diversity of the data they seek, the interpretation of such data is best considered within the context of each of the methods. The analysis and interpretation of historical data, for example, is best viewed in the light of the historical method, its objectives, and its limitations. For the present it is important to note that, regardless of the adequacy of the data and of the procedures by which they are processed, data do not interpret themselves, and that it is the investigator who must pass judgment on their meaning from the standpoint of the problem under investigation.

It is also essential to recognize that errors can be made in interpretation—just as they can in any of the other steps of the scientific method—and the specific errors to be guarded against vary with the different research methods. The following are among the more common errors of interpretation:

1. Failure to see the problem in the perspective of its theoretical and empirical setting, perhaps as a result of an inadequate grasp of the problem in its broad sense and too close a focus

on its immediate aspects. Thus, the Hartshorne and May studies are not to be interpreted as supporting the view that human behavior is inconsistent and haphazard, but rather that the consistency is internal rather than external.[9]

2. Failure to appreciate the relevance of the various elements of the situation, resulting from such factors as an inadequate grasp of the problem, too rigid a mind-set, or even a lack of imagination. This may cause the investigator to overlook the operation of significant factors—for example, motivation and teacher competence in studies of the effectiveness of teaching methods, or selective migration and test fairness in studies of regional or class differences in intellectual ability. Consequently, the outcomes of the study are attributed to the wrong antecedent. A parallel error is the failure to see crucial relationships to be pursued, and the resulting failure to obtain data vital to the investigation.
3. Failure to recognize limitations in the research evidence—such as non-representativeness in sampling, biases almost inevitable in the data concerning certain phenomena, and inadequacies in the research design, the data-gathering instruments, and/or the statistical analysis. Particularly incapacitating from the standpoint of the study is the common failure on the part of the investigator to see that the research design could not possibly lead to any other results than those that were obtained. Thus, the interviewing of students or parents regarding their attitude toward the school is almost sure to lead to endorsement. Similarly, it is said that personality is a more important consideration in teacher effectiveness than is knowledge of subject. Inasmuch as teachers cannot be certified to teach unless they know their subject well enough to pass the required college courses, the operation of the factor of knowledge of subject-matter is restricted sufficiently to give precedence to other, more unlimited factors in the situation. (A similar error, noted by Russell in connection with the classic studies in learning, is reported in Chapter 12).[10]

Of a parallel nature is failure on the part of the investigator to make the relative limitations of his study sufficiently explicit so that while the study is correct, it is misleading in that it promotes misinterpretations and/or over-extension of its findings and conclusions. A similar error can be promoted by

[9] See Chapter 15.
[10] Bertrand Russell, *Philosophy* (New York: W. W. Norton, 1927), p. 29–30.

failure to report the study in sufficient detail to permit the reader to gain an adequate grasp of its nature.

SYNTHESIS AND ORIENTATION

The discussion of the steps of the scientific method presented in this chapter has been restricted to an overview designed to provide continuity and to bring out the unity of educational research. Inasmuch as the implementation of the scientific method is relatively specific to the particular type of research in which it functions, it seems more appropriate to integrate the treatment of the more specific and detailed aspects in the context of the presentation of the various types of research which follow.

On the other hand, what is significant is not the peculiarities of the application of the scientific method to the specific situation, but rather its universality. What needs to be stressed is that science cuts across the arbitrary lines that separate the various disciplines, and that scientists, regardless of disciplinary allegiances, subscribe to a common core of procedures and attitudes in their search for truth. That the general level of sophistication at which the various disciplines operate should vary is inescapable in view of the degree of relative development of each and the complexity of the material with which they deal, but the co-ordination of their efforts toward a common objective to be attained by subscription to a common method makes the difference one of degree rather than of kind.

In this connection, Hillway[11] presents the role of the investigator as that of a detective. Developing this parallel, he points out that the scientist must be alert and trained to seek clues that will develop into fruitful hypotheses, that he must be familiar with sources of information, and that he must be able to extract the desired information quickly and effectively. Like his detective counterpart, the scientist must not solve his problem on the basis of opinions, no matter how logical they may appear. While he will have to start with hunches and opinions, these are only hypotheses which he must check for validity. He must evaluate all the information he gathers before attempting to synthesize it with respect to his hypothe-

[11] Tyrus Hillway, *Introduction to Research* (Boston: Houghton-Mifflin, 1956), p. 57ff.

ses. Finally, both the scientist and the detective must test their hypotheses against objective evidence, not mere plausibility. Hillway points out further that, just as detectives differ in their ability to sense important clues and to develop them through logical and empirical considerations, so do investigators differ in their ability to derive fruitful hypotheses and to develop them through the accumulation of the data collected in line with those hypotheses.

SUMMARY

1. Even though we depend so critically on science for our material and social welfare and progress, most people have only an inadequate conception of its nature and purpose. We are especially lacking in appreciation of the role of systematic research in the development of the social sciences.

2. The wise selection of a research topic is among the most crucial aspects of success in research. Unfortunately, the student selecting a problem for thesis or dissertation purposes is generally restricted by a lack in the areas of problem-consciousness, knowledge and perspective of the field, research and statistical competence, and access to data.

3. The question of its possible contribution and its feasibility are among the more important criteria for the selection of a research problem. There is, however, no standard set of rules that will provide the student with a suitable problem. Familiarity with the field and imagination are among the more important attributes facilitating the wise selection of a research topic.

4. If a problem is to serve its function as a guide in the planning and the conduct of a research study, it must be clearly delineated. It must strike a balance between excessive scope—and resulting unmanageability—and overstriction, with its consequent artificiality.

5. Whenever possible, the problem should be converted into a hypothesis to be tested, for hypotheses highlight the direction in which the study is to go, the data that need to be collected in its verification, and the way these are to be processed to provide an adequate answer. Not only does a hypothesis alert the investigator to relevant aspects of the situation and permit him to refine his research design, but it also provides him with the framework for the interpretation of the findings and the derivation of conclusions. Generally, the formulation of a hypothesis goes hand in hand with the selection and clarification of the problem. Imagination and familiarity with the field—as well as persistence and a critical attitude—are important factors in the formulation of a good hypothesis.

6. The most important criterion of a good hypothesis is its

testability. Are its implications, when stated in operational terms, compatible with known facts, and, further, are they compatible at the empirical level with the results of research specifically designed to test their validity? A hypothesis is never proved; it is simply sustained or rejected, and, like a theory, a hypothesis may be useful even though it is partially in error. On the other hand, if its significance and scope warrant it, a hypothesis that is sustained may eventually attain the status of law or principle.

7. The research design must be amenable to providing data on the basis of which the problem can be resolved. Inasmuch as no study can be more adequate than the data on which it is based, competence in research requires familiarity with the principles of tests and measurements—particularly with the concept of validity. The researcher must also be familiar with statistical procedures capable of the adequate analysis of the data that have been collected.

8. Among the more common errors in the interpretation of the results of research are failing to see the significance of the data, failing to see the limitations of the research design, overlooking contrary evidence, mistaking coincidence for cause-and-effect, and reversing the cause and the effect. The best safeguards against such errors are common sense and insight into the field.

9. Each of the different methods of educational research presents special problems. However, what is significant is not the unique nature of the different methods but rather the universality of the scientific principles that underlie the various approaches necessary to deal with the varied problems encountered in a field as broad and scientifically undeveloped as education.

PROJECTS and QUESTIONS

1. The selection of a problem is always among the most difficult tasks facing the graduate student.

    a) List a number of broad general areas in the field of education. What is the present research status of each? Which are in need of further investigation? Identify one or more researchable problems.

    b) For the problems above, elaborate on (1) their practical and theoretical significance; (2) their amenability to research; (3) the obstacles in the path of their solution.

    c) Plan a research design for the investigation of one of the problems above and present the design for class discussion and evaluation.

2. Discuss specific ways in which the graduate student might locate a suitable topic.
3. Discuss "brainstorming" as a means of getting new ideas for research purposes.
4. "Unfortunately, some writers make their facts conform to their

hypotheses rather than vice versa." (Brickman: *A Guide to Research in History*. p. 116) Evaluate the above statement. What safeguard might be taken to prevent this from occurring? Would refraining from starting from a hypothesis be the answer?

SELECTED REFERENCES

ALILUNAS, LEO J. "Needed Research in Teacher Mental Hygiene," *Journal of Educational Research,* 38 (May 1945) : 653–65.

BETTS, EMMETT A. "Unsolved Problems in Reading," *Elementary English,* 31 (October 1954) : 325–9.

BUSWELL, GUY T. "Needed Research in Arithmetic," in *The Teaching of Arithmetic.* 50th Yearbook, National Society for the Study of Education, Pt. 2. Chicago: University of Chicago Press. 1951.

CHAMBERLIN, THOMAS C. "The Method of Multiple Working Hypotheses," *Scientific Monthly,* 59 (November 1944) : 357–62.

CLYMER, THEODORE W. "Some Current Needs in Educational Research," *Phi Delta Kappan,* 40 (March 1959) : 253–7.

COLADARCI, ARTHUR P., *et al.* "Research Priorities in Educational Administration," *Journal of Educational Research,* 47 (April 1954) : 625–30.

COOMBS, CLYDE H. "Theory and Methods of Social Measurement," in Festinger, Leon and Katz, Daniel (eds.) . *Research Methods in the Behavioral Sciences.* New York: Dryden, 1953.

DEWEY, JOHN. *How We Think.* Boston: Heath, 1933.

FOX, ROBERT S. "Needed Research in the Elementary School Curriculum," *University of Michigan School of Education Bulletin,* 23 (1951) : 39–42.

FRASER, THOMAS P., *et al.* "Needed Research in Science Education: Symposium," *Science Education,* 40 (December 1956) : 363–77.

GOOD, CARTER V. "Criteria for Selection of the Research Problem," *Peabody Journal of Education,* 19 (March 1942) : 242–56.

GRAY, WILLIAM S. "Needed Research in Reading," *Elementary English,* 29 (February 1952) : 100–8.

HALL, ROY M. "Research Needed in Education," *School Life,* 42 (March 1960) : 6–8.

HENDRICKSON, GORDON. "Some Needed Research in Elementary Education," *Elementary School Journal,* 51 (November 1950) : 127–35.

MAASKE, ROBEN J., *et al. Needed Research in Teacher Education.* Study Series, No. 2. Oneonta: American Association of Colleges for Teacher Education, 1954.

MCCONNELL, THOMAS R. "Some Unresolved Problems in Secondary Education," *North Central Association Quarterly,* 27 (January 1953) : 258–66.

MCGREGOR, DOUGLAS. "Scientific Measurement and Psychology," *Psychological Review,* 42 (May 1935) : 246–66.

MADGE, JOHN. *The Tools of Social Science.* New York: Longmans, Green, 1953.

MALLINSON, GEORGE C. "The Five Most Needed Research Investigations in the Teaching of Science and the Teaching of Mathematics," *School Science and Mathematics,* 54 (June 1954): 428–30.

NEWLAND, T. ERNEST. "Essential Research Directions on the Gifted," *Exceptional Children,* 21 (May 1955): 292–6.

O'SHEA, HARRIET E. (Chairman). "Needed Research on Gifted Children," *American Psychologist,* 9 (January 1954): 77–8.

PRICE, HENRY V. "Needed Research in the Teaching of Science and Mathematics; Report of the Projects and Research Committee," *School Science and Mathematics,* 57 (May 1957): 389–90.

RATHS, LOUIS E. "The Role of Evaluation in Research Design," *Educational Leadership,* 13 (April 1956): 412–4.

ROSE, ARNOLD M. "The Selection of Problems for Research," *American Journal of Sociology,* 54 (November 1948): 219–27.

SCATES, DOUGLAS E. "Types of Assumptions in Educational Research," *Journal of Educational Psychology,* 25 (1935): 350–66.

SHANE, HAROLD C. and ROVNER, SEYMOUR. *A Selected Annotated Bibliography of Evaluation Instruments and Related Materials.* Evanston: School of Education, Northwestern University, 1950.

VAN DALEN, DEOBOLD B. "Role of Hypotheses in Educational Research," *Educational Administration and Supervision,* 42 (December 1956): 457–60.

WALLER, WILLARD. "Insight and Scientific Method," *American Journal of Sociology,* 40 (November 1934): 285–97.

WOLFLE, DAEL L. "Fetish of an Experiment," *Science,* 125 (February 1957): 177.

ZIRBES, LAURA. "Needed Research in Education," *Educational Leadership,* 10 (November 1952): 129–31.

PART II: RESEARCH TECHNIQUES

If we are stupid, we are stupid by choice—for well within walking distance of any of us stands a library with its unlimited knowledge and unlimited wisdom. ANONYMOUS

5 The Library

Man is the only animal that does not have to begin anew in every generation, but can take advantage of the knowledge which has accumulated through the centuries. This fact is of particular importance in research which, as we have seen, operates as a continuous function of ever-closer approximation to the truth. The investigator can be sure that his problem does not exist in a vacuum, and that considerable work has already been done on problems which are directly related to his proposed investigation. The success of his efforts will depend in no small measure on the extent to which he capitalizes on the advances—both empirical and theoretical—made by previous researchers.

THE REVIEW OF THE LITERATURE

An essential aspect of a research project is the review of the related literature. Such a review represents the third step of the scientific method outlined by Dewey and other educational philosophers, and the serious student of research will find an exhaustive survey of what has already been done on his problem an indispensable step in its solution. The survey of the literature is a crucial aspect of the planning of the study, and the time spent in such a survey invariably is a wise investment. Students frequently fail to appreciate the importance of the review of the literature; they are likely to feel they know enough about their problem and that their task is to get on with

its solution. This feeling is frequently reflected in a tendency to dismiss the task as completed after a few articles have been reviewed and, especially, in the fact that the relevant literature, even when thoroughly reviewed, is often inadequately integrated with the rest of the paper.

The review of the literature is an exacting task, calling for a deep insight and clear perspective of the overall field. It is a crucial step which invariably minimizes the risk of dead-ends, rejected topics, rejected studies, wasted effort, trial-and-error activity oriented toward approaches already discarded by previous investigators, and—even more important—erroneous findings based on a faulty research design. The review of the literature promotes a greater understanding of the problem and its crucial aspects and ensures the avoidance of unnecessary duplication. It also provides comparative data on the basis of which to evaluate and interpret the significance of one's findings. In addition, it contributes to the scholarship of the investigator.

The published literature is a fruitful source of hypotheses. Not only does it present suggestions made by previous investigators and writers concerning problems in need of investigation and hypotheses in need of testing, but it also stimulates the research worker to devise hypotheses of his own. As he reacts to the designs, findings, and conclusions of other investigators, he can get insights which he can incorporate into an improved research design. Capitalizing on the successes and errors of others is certainly a more intelligent approach to a problem —especially one as broad as a thesis or dissertation—than is imagining that one is born equipped with a radar system that will guide him unerringly on target, and at the same time guard him against all pitfalls. Rarely does the neophyte have such insight into his problem that he cannot profit from the work of others; no experienced researcher would think of undertaking a study without acquainting himself with the contributions of previous investigators.

THE LIBRARY

The Organization of the Library

The library is the storehouse of the knowledge and wisdom which has accumulated since the beginning of time, for

in a general sense, whatever is worth knowing is probably recorded in one of the volumes in the library. Until relatively recently, man's progress was seriously hampered by the lack of source material. Until the invention of printing, and even for decades after, books were available only to the rich, and arriving at knowledge beyond that of personal experience was a relatively difficult task. Today, in contrast, we have such an abundance of written material that it is almost impossible to keep abreast even of one's own specialty. We are now in the age of specialist and the abstract; our only barriers to knowledge and wisdom are time, motivation, and intellectual endowment.

Because of the library's tremendous assortment of material—of endless quantity, variety, and complexity—it is imperative that the researcher know how to locate and to use what is available, for without such a skill, he is simply a hunter lost in the forest. Effective research presupposes a good grasp of the organization of the library and of its content.

Library work is a complex science. The wide variety of materials found in any library calls for a highly organized and involved system of classification. It is not expected that the graduate student will attain the proficiency of a librarian, but facility in the use of the library and its materials is, in a sense, the key to graduate studies. Although he occasionally will have to consult a professional librarian in order to locate unusual material, it is absolutely necessary that the graduate student be able to locate the more common sources quickly and efficiently, and that he be able to extract the information contained therein with dispatch and accuracy—relying on librarians for help only in special cases.

Fortunately, except for the physical aspects, the plan of organization is the same from library to library, and proficiency in the use of one library can be transferred to another. Furthermore, the many indexes and guides which are so fundamental in locating material are, of course, the same throughout America and, generally, throughout the English-speaking world.

The library is big: the library on the campus of a large university may have a million bound volumes specially cataloged, and in addition may subscribe to some 5,000 serials. Through loans the library can gain access to over a million

articles published in professional journals in any year. (The Library of Congress in 1961 reported total holdings of over 41 million separate items.) Except in the smallest libraries, holdings are departmentalized—the circulation department, the reference room, the periodicals room, the government publications, the stack area, the reading rooms, and so on—each designed to provide a special service with an overall maximum efficiency. Many of the larger libraries have special graduate-seminar rooms to which library books may be delivered for short periods of time; nearly all have carrels in the stack area where graduate students and faculty members can work close to the books they are likely to need. All of this can be rather complicated—in fact, hopelessly complicated—for the person who does not understand the system of organization. This, the graduate student needs to learn, and though it is hoped he has had contact with the library as an undergraduate, it must be recognized that his needs as a graduate student are much more complex than they were then.

Proficiency in the use of library—and thus in the review of the literature—consists of the ability to locate sources directly, to browse through multiple sources quickly, to cull relevant material, and to interpret and organize what one has accumulated. The specific procedures by which this can be done can be presented in further detail.

1. The logical starting point is to get a clear picture of the problem to be solved. Without this perspective, the review of the literature is a matter of reading at random, hoping that a problem will emerge. It is generally advisable to get first an overall view by consulting a general source, such as a textbook, which is more likely to give a coherent picture of the field than is a more specialized source. A textbook is also more likely to deal with the theoretical aspects of the problem, and thus provide the prospective investigator with an overall framework within which his problem and its many aspects can be seen in perspective.

2. Having grasped the general nature of his problem, the investigator should orient himself toward the empirical research done in the broad area in which his problem lies. The best reference for this phase is the *Encyclopedia of Educational Research*, and the *Review of Educational Research* for more up-to-date findings. An education society yearbook in the area is another ideal source. The student's major concern at this point should be

to get a clear picture of the field as a whole; specific details are important only after he has attained the structure into which they can fit meaningfully. Sources should not be read for their own sake, but rather for what they can contribute to rounding out a pattern which appears logical from the investigator's present view of his problem. This is a crucial consideration from the standpoint of effective library research. The investigator must be so familiar with his problem that he can judge the relevance of his readings. In fact, he should operate from a topical outline and a tentative set of classifications, so that whatever he reads can be immediately filed rather than merely accumulated. He must also know what to look for in order to gain full perspective of his problem, so that any lack in his search to date becomes immediately noticeable.

3. Effective library work also depends on the ability to read at a high rate of speed. The student must learn to skim material to see what it has to contribute to the study; only after its relevance has been established should it be read in detail. Whatever is not pertinent to the study, regardless of its personal appeal, should be simply noted for later referral and dropped. Exploring all kinds of side issues merely sidetracks the investigation. Surveying the literature for the purpose of conducting research is not just "a pleasant excursion in the wonderful world of books," it is a precise and exacting task of locating specific information for a specific purpose. Any tendency to wander should be curtailed.

4. The search for library material must be systematic and thorough. The investigator generally should begin by collecting his reference cards, for unless the bibliography is developed systematically useful sources may be overlooked. In locating references from the *Education Index,* for example, it is generally desirable to work backward from the current volume. Judging the utility or the futility of an article by its title is always precarious; generally it is best to record anything that may be useful, and then to rely on one's ability to skim in order to save time and yet not overlook significant studies.

When a large number of references are to be copied, they should be typed, if possible; handwriting tends to be slow and is often illegible from the standpoint of the precision required here. Better still: Why not thermofax pages of the *Index* in which a number of pertinent entries appear? These copies can then be used directly in the search for material or in typing out a set of cards. It is generally best to collect the bulk of the refer-

ences at one time, so that the cards can be sorted and duplicates eliminated. Of course, other references will be found in the bibliographies of the articles read.

5. Notes should be taken systematically in the light of such criteria as uniformity, accuracy, and ease of assembly. Each entry should be separate; references should be recorded, one to a card, with complete bibliographic data entered on one side of the card. The content can be recorded below, or possibly on the reverse side. Consistency is important—a note on the back of a card may be overlooked, if one does not generally put notes there. Each note must be carefully labeled: nothing is more frustrating when it comes to writing than to find a note which is not clear as to why it was collected, where it came from, or what it is supposed to mean.

6. The investigator should take as complete notes as he might need. On the other hand, taking unnecessary notes is wasteful and, though it is better to err on the side of too much rather than too little, there is no substitute for knowing precisely what is useful and what is useless. While this ideal is never attained, the adequate researcher strikes a close balance between keeping notes to a minimum while, at the same time, also keeping to a minimum the need to check a source a second time because of failure to take adequate notes on first contact. It is also better not to recopy or to fill in details later: recopying invites errors, and memory is invariably bad at the end of a full day of library work, when so much material of a relatively similar nature from relatively similar sources has been gathered.

It is essential that a general evaluation of each source be made, rather than simply a summary of its contents. Such an evaluation is necessary both in presenting the study in the section on the review of the literature, and in using the study as background for the interpretation of the findings of the present investigation.

7. The actual note-taking process is always a chore. Long hours spent taking notes by hand can be torture. Too frequently, tediousness leads to impatience, to carelessness and illegibility, and to a tendency to cut corners and rely on one's memory in a misguided attempt to expedite the process. As a result, the final product is frequently short of ideal; at worst, it may be conducive to serious error.

The usual procedure for recording references is to take notes directly on 3 x 5 or 4 x 6 cards, labeling each for topic or topics. The author has found the IBM card superior to either

size index card; it is thinner and of a better grade of material; it can be obtained with different color stripping at the top to identify topics; it can even be punched on any number of classifications and sorted by machine. It is also cheaper.

Probably the biggest stumbling block to effective library research is the tediousness of handwriting. One alternative is typing. Most libraries have typing rooms for the use of graduate students and faculty members. If typing facilities are available, half-sheets or even whole sheets of paper—with a carbon as a record—which can then be cut, sorted, pasted, and otherwise manipulated to expedite the writing of the first draft of the report, should be used. Another very satisfactory procedure is to dictate notes directly from the references into a portable tape recorder for transcription at one's convenience; this method is both simple and efficient.

The student should take advantage of modern facilities wherever possible. He should never copy tables, for example, but should have them thermofaxed so that he can have an authentic copy of the original when he is writing his report. Passages that may be used in a quotation should be reproduced rather than copied to preclude the risk of errors. Most libraries have duplicating facilities available for a very nominal fee. It also should be pointed out that quotations—and even other material—should never be taken from a secondary source, except as a last resort. The more removed from the original source the data are, the greater the risk of error.

The Card Catalog

The holdings of the library fall under two major classifications: books and periodicals—and a third category of miscellany which includes government documents, manuscripts, pamphlets, references materials, maps, and so on. The vast bulk of the library's collection is cataloged under *books* which includes books, booklets, yearbooks, pamphlets, and certain serials, most of which can be obtained on loan from the circulation department. Each volume in this category ordinarily covers a single topic which can be identified by its title, and a reader interested in a given volume generally would want to read all, or a good part, of its contents.

In contrast, periodicals usually cover a wide variety of topics, and the reader is interested not in a whole volume but in particular articles which must be traced through the use of an

index. Generally, periodicals are not coded but are filed alphabetically in open stacks and cannot be taken out of the library. In addition, references—dictionaries, encyclopedias, and guides of various sorts—are housed in open shelves in the Reference Room. These are cataloged as regular bound volumes held by the library, but their circulation generally is limited to "room use."

In most university libraries, all students have ready access to the open shelves of the Periodical and Reference Rooms, but only graduate students have access to the stacks. This special privilege is partially in recognition of the greater dependability of the graduate student, and partially in recognition of the fact that, in order to do research, graduate students must have the opportunity to browse through numerous sources quickly.

No matter how the student gets his books, however, he must be able to identify what he wants. This is done through the card catalog which generally lists all library holdings, except the periodicals and government documents. Although the Reference Room and the special libraries on campus have duplicate catalogs for their holdings, the card catalog of the Circulation Department is the master list of all material cataloged in any branch of the library, and any source can be traced from this catalog.

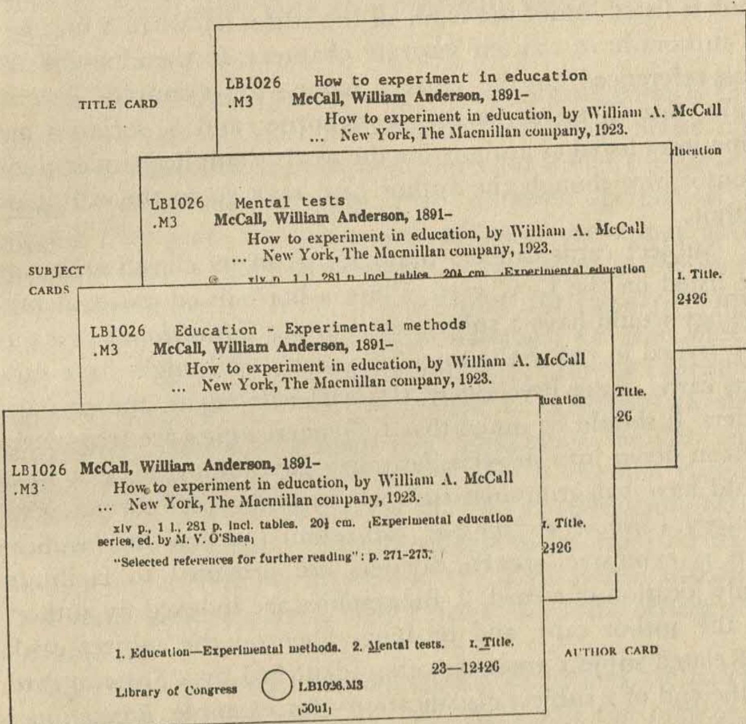
The primary purpose of the card catalog is to record what is contained in the library. In a sense, it is an index to the library. As its name implies, it is a listing of the library's holdings in various areas, describing each item briefly, and giving the classification code so that it can be readily and correctly identified, and so that it can be shelved in a way that it can be located with a minimum of delay and a maximum of certainty.

Each volume is cataloged under author, title, and subject on separate 3 x 5 cards, arranged in alphabetical sequence in row after row of drawers. All cards pertaining to a given volume list essentially the same information, but they are filed differently. Thus John Doe's *Nuclear Physics* would be filed under *Doe* for the author card, under *Nuclear* for the title card, and under *Physics* for the subject card.

Both the general format of a catalog card, and the distinction between author, subject, and title card can be seen

from the accompanying illustration. Each card lists the following basic information:

1. the name of the author;
2. the title of the book;
3. the imprint (edition, place, publisher, and date of publication);
4. special information (number of pages, preface, bibliography, size, illustration, number of volumes);



5. the subject classification (s) and other entries under which separate cards are to be found;
6. the Library of Congress and the Dewey Decimal call number; and
7. the L. C. (Library of Congress) card number.

In addition, each library types its own call number on the top left-hand corner of each card. If the library uses either the L. C. or the Dewey Decimal system, this may be the same num-

ber as that already printed on the card, or it may be whatever call number the volume carries in the particular library's classification system. Occasionally, a card will carry additional information—for example, a listing of the major parts of a volume.

The author card is the master card and, on occasion, it may be the only card in the card catalog for a given volume, but a library would never fail to have an author card for each volume. Some of the peculiarities of the card catalog are: 1. separate cards are made for co-authors of a book, but the book is listed under the name of the editor (s) when a number of authors have written separate chapters; 2. pseudonyms are cross-referenced to the correct name—for example, Twain, Mark, *See* Clemens, Samuel Langhorne; and 3. societies are sometimes listed as authors for the works compiled under their sponsorship, though the author card may go to the editor or author.

Subject cards can be found for as many classifications as are listed on the L. C. card. Thus, a book in educational psychology would have a subject card under *Educational Psychology* (typed in red at the top of the card), and might have subjects cards under *Psychology, Child Development,* and perhaps others. It should be noted that 1. General topics are frequently broken down into sub-classifications: *Psychology,* for example, could have sub-groupings such as *Psychology, Abnormal; Psychology, Clinical; Psychology, Educational.* When the classification is extensive, special dividers are provided to facilitate ready location of a card. 2. Biographies are indexed by author, on the author card, and by biographee on the subject card. 3. Related subject areas are often identified by a *See also* card at the end of a subject classification—for example, *Educational Psychology, See also Child Development.*

The title card carries the title of the volume, typed directly above the name of the author. It is filed alphabetically according to the first major word of the title. Some volumes do not have a sufficiently distinctive title—for example, *Introduction to the Study of History*—to warrant separate listing and therefore do not have a title card.

All cards are filed alphabetically with guide cards at intervals. Cards *about* an author come after cards *by* him—that is, author cards take precedence over subject and title cards. Com-

plete words are filed before words of which they constitute only a part—for example, New York is filed before Newman. Hyphenated words like *pre-election* are filed without regard to the hyphen. Names beginning with Mc are filed as if they were spelled Mac, and all abbreviations are filed as if they were spelled out in full—for example, Saint, Mister, United States, and so on. Titles beginning with a numeral are filed as though the number were spelled out—for example, *The 100 Best Sellers* would be filed under *One Hundred Best Sellers.*

Classification Systems

The crucial aspect underlying library organization is the classification system. Most libraries in the United States operate either under the Library of Congress or under the Dewey Decimal system, though some small libraries operate on a system of their own, and some of the older libraries may have a modified system. Undoubtedly, the L. C. system is the most flexible and comprehensive: it is designed for classifying unlimited quantities and varieties of material. On the other hand, it is a relatively new system and many libraries, already on a system of their own at the time the L. C. system was devised, did not elect to reclassify their holdings to the new system. Many libraries operating on the Dewey Decimal system have found it adequate for their needs and are continuing with it; others have devised certain modifications in order to take advantage of some of the features of the L. C. system without at the same time incurring the expense of reclassifying all their holdings.

Both the L. C. and the Dewey Decimal systems are based on the allocation of a code for each field and the breaking down of these fields into finer and finer subclassifications. The major classifications of the two systems are shown below.

L. C. Classification

A. General Works
B. Philosophy, Religion
C. History
D. World History
EF. American History
G. Geography, Anthropology
H. Social Sciences
I. Vacant

Dewey Decimal Classification

000 General References
100 Philosophy, Psychology
200 Religion
300 Social Sciences
  310 Statistics
  320 Political Science
  330 Economics
  340 Law

J. Political Science
K. Law
L. Education (general)
LA. History of Education
LB. Theory of Education
LC. Special Forms and Applications
LD U.S. Schools
LE. American Education (Outside U.S.)
LF. Education—Europe
LG. Education—Asia, Africa, Oceania
LH. College and School Magazines, student periodicals
LJ. Fraternities and their Publications
LT. Textbooks
R. Medicine
S. Agriculture
T. Technology
U. Military Science
V. Naval Science
W. vacant
X. vacant
Y. vacant
Z. Library Science

350 Administration
360 Welfare Associations and Institutions
370 Education (general)
370.1 Theory and Philosophy of Education
370.9 History of Education
371 Teachers—Methods, Discipline
372 Elementary Education
373 Secondary Education
374 Adult Education
375 Curriculum
376 Education of Women
377 Religion, ethical Education
378 Higher Education
379 Public School, relation of state to education
380 Commerce, Communication
390 Customs, Costumes, Folklore
400 Philology
500 Natural Science
600 Useful Arts
700 Fine Arts
800 Literature
900 History

Inter-Library Loans

Materials not available in one library can often be obtained from another on loan. The loan is always from the lender library to the library at which the student is enrolled and then to him. Such loans are simply a courtesy, not an obligation[1] and the conditions and mechanics under which such loans can be effected are governed by American Library Association regulations. Thus, loans are limited to material not readily available through purchase, and they generally do not cover such items as irreplaceable manuscripts. The Library of Congress itself has certain facilities for honoring requests that conform to its loan policies. Generally, however, loans are effected on a regional basis. Most loans are for two weeks, and the borrower usually pays transportation charges both ways.

[1] It should be noted further that a library is in no way obligated to provide the student with facilities it does not have.

The student should not overlook the possibility of purchasing some of the material he needs. Where extended use is indicated, the small cost may be more than repaid by a saving in time and convenience. This is particularly true of such materials as pamphlets and special reports. If a journal article is needed, it can save time and even money to have a thermofax copy made or even to request a typed copy. Most libraries can provide this kind of service. In fact, even a sizable volume can be microfilmed. Finally, when a certain amount of browsing needs to be done, the student should plan to visit a neighboring library where the needed volumes are available. Most libraries have the *National Union Catalog,* either in photo-print or card form, of the complete holdings of the Library of Congress. *The Union List of Serials in Libraries in the United States and Canada* (New York: Wilson, 1962) will identify the major libraries having certain serials. If necessary, a student can write to a library asking if it has the publications he needs.

LOCATING LIBRARY SOURCES

The Periodical Literature

The other major component of the library collection is the periodicals. This section is of particular interest to the research worker since it is here that he will find most of the material needed for his review of the literature. The use of periodicals is predicated on an entirely different basis than that which governs the circulation material. Here we are interested not in a given volume, but in isolated articles which must be traced individually. And, though certain journals tend to have articles related to certain topics, the effective use of journal material requires the use of a suitable index, the most important of which, as it applies to educational materials, is the *Education Index* (New York: Wilson, 1929–date). In fact, at times it may be necessary to use a number of indexes and guides—and this may involve complexities, many of which are beyond the scope of this text. The discussion will be simply suggestive, and the student is advised to consult other sources such as Alexander and Burke, *How to Locate Educational Data and Information* (4th ed.; New York: Teachers College, Columbia University, 1958).

The *Education Index* is undoubtedly the most important library tool of the worker in educational research. Issued ten times a year from September through June (and cumulated in annual and two-year volumes), the *Index* lists practically all current references to educational literature—periodicals, books, and pamphlets. Specifically, it indexes and cross-references by author and subject: 1. the contents of nearly 200 educational periodicals as listed inside the front cover of every volume; 2. all references and books in professional education; 3. all National Education Association (N.E.A.) publications and many of the publications of other professional societies in education and related fields; 4. all United States Office of Education documents; 5. published bibliographies and book reviews; and 6. all educational articles indexed in the various Wilson indexes.

Such a large number of other indexes, abstracts, and guides are available to the research worker that only a few can be mentioned here.

References to periodicals include:

1. *Ulrich's Periodical Directory* (Eileen C. Graves, ed., 10th ed.; New York: Bowker, 1962). Covers a selected list of current periodicals, arranged by subject.
2. *American Educational Press Yearbook* (New York: American Education Press, 1926–date). Gives rather complete information on all American periodicals dealing with education.
3. *Union List of Serials in Libraries in the United States and Canada.* (See above.)
4. *Ayer's Directory of Newspapers and Periodicals* (Philadelphia: Ayer & Son, 1880–date).
5. *Subject Index to Periodicals* (London: London Library Association, 1915–date).

The more common general indexes are:

1. *Readers' Guide to Periodical Literature* (New York: Wilson, 1900–date). Indexes all general sources. It covered education to 1929, when the *Education Index* took over that area. *Readers' Guide* is the successor to *Poole's Index to Periodical Literature* (Boston: Houghton-Mifflin, 1802–1881) and *19th Century Guide to Periodical Literature* (New York: Wilson, 1890–1899). There is also an *Abridged Readers' Guide to Periodical Literature* (New York: Wilson, 1935–date) which covers 35 periodicals.

2. *International Index* (New York: Wilson, 1913–date). Indexes a somewhat more selective list of periodicals in the social sciences, including a number published abroad.
3. *Ireland's Index to Indexes* (Boston: Faxon, 1942).
4. *Vertical File Index* (New York: Wilson, 1932–date). An excellent source of low-cost materials.
5. *Facts on File* (New York: Facts on File, Inc., 1940–date). Indexes current events semi-monthly, monthly, quarterly, and annually.
6. *Public Affairs Information Service Bulletin* (New York: Public Affairs Information Service, 1915–date).
7. *New York Times Index* (New York: *New York Times*, 1913–date). Indexes the contents of the *New York Times* by subject, title, person, and organization.

The best known indexes and guides covering the periodical literature of a professional nature (besides the *Education Index*) include:

1. *The Catholic Periodical Index* (New York: Catholic Library Association, 1930–date). Indexes the periodical literature dealing with Catholic education.
2. *Psychological Abstracts* (Washington, D.C.: American Psychological Association, 1927–date). Abstracts the psychological literature. It is published monthly and indexed yearly. *Psychological Abstracts* supersedes *Psychological Index* (1895–1936), which carried a bibliography of the international literature on Psychology from 1894 to 1928.
3. *Child Development Abstracts and Bibliography* (Washington, D.C.: National Research Council, 1927–date). Published monthly and indexed yearly.
4. *Education Abstracts* (Bloomington: Phi Delta Kappa, 1936–44); *Loyola Education Digest* (Chicago: Loyola University, 1924–43), and *Loyola Education Index* (Chicago: Loyola University, 1928). All have discontinued publication.
5. *Educational Abstracts* (UNESCO, Paris: Education Clearing House, 1949–date). Covers world scene in education.
6. *Education in Lay Magazines* (Educational Research Service, Washington, D.C.: The Service, N.E.A., 1944–date).
7. *Record of Current Educational Publications* (Washington, D.C.: United States Office of Education, 1912–32). Contained an annotated listing of educational publications and is generally considered the best predecessor of the *Education Index*.
8. *Free and Inexpensive Learning Materials* (10th ed.; Nashville: Peabody College for Teachers, 1960).

9. *Filmstrip Guide* (New York: Wilson, 1948–date) and *Educational Filmstrip Guide* (New York: Wilson, 1948–date).
10. *Guidance Index* (Chicago: Science Research Associates, 1938–date) and *Occupational Index* (New York: Occupational Index, Inc., New York University, 1943). The latter indexes over 100 periodicals in the area of occupations by author, title, and subject. It is published monthly and cumulated yearly.

Similar indexes and guides are found in specialized areas of education, for example, business education, agriculture education, art education, and music education. Those journals which are not indexed systematically in the regular indexes must be scanned individually through their table of contents.

Books and Textbook Materials

Probably because of their individual importance, regular books are most systematically and extensively covered by bibliographic services. In fact, any book published in this country, or any of the English-speaking countries, generally can be located if one knows the author, the title, or even the approximate date of publication.

The most useful list of books published in the English language is the *Cumulative Book Index* (New York: Wilson, 1898–date) which is issued monthly, with semi-annual and annual cumulations and larger volumes covering two-year and four-year periods. These cumulations make available, in one alphabet, a world list of books in the English language, citing publisher, edition, date of publication, price, paging, L. C. card number, and so on.

Other very useful guides to books include:

1. *Publishers' Trade List Annual* (New York, Bowker, 1873–date). Covers the catalogs of over one thousand publishers.
2. *Books in Print* (15th ed.; New York: Bowker, 1962, 1873–date), and *Textbooks in Print* (New York: Bowker, 1870–date). Both of these index, by author and title, appropriate books from (1) above.
3. *Subject Guide to Books in Print* (New York: Bowker, 1957–date). The 1962 edition lists 154,000 titles under 24,000 subject headings and 35,000 cross references.
4. *Subject Collections* (Lee Ash, compiler, 2nd ed.; New York: Bowker, 1961).

5. *Book Review Digest* (New York: Wilson, 1905–date) and *American Library Association Booklist and Subscription Books Bulletin* (Chicago: American Library Association, 1955). Both of these not only list books shortly after their publication, but also provide a review of their worth. The monthly issues of *Book Review Digest* furnish a cross-section of reviews of current literature. The latter also contains appraisals of encyclopedias, dictionaries, and other references.
6. Book review and listings are frequently found in the *Books Received* and *Book Reviews* sections of certain journals. They can be located through the indexes of these journals. *Contemporary Psychology* (Washington, D.C.: American Psychological Association, 1956–date) is devoted almost exclusively to reviews of books in Psychology.

Most professional societies provide listings of their own publications (in addition to coverage in the *Education Index*). The National Education Association publishes an annual catalog of the publications of the parent organization and its member groups. The American Council on Education has available on request a complete listing of its publications since 1918 (some are out of print). The National Society for the Study of Education has published a yearbook in two parts yearly since 1901. The Association for Supervision and Curriculum Development also publishes a yearbook (1944–date).

The publications of such organizations as The National Society for the Study of Education and the Association for Supervision and Curriculum Development which are issued on a recurrent basis are listed in the card catalog, the *Education Index*, the *Cumulative Book Index*, and *Publishers' Weekly*. A complete listing of educational associations and their publications of a systematic nature can be found in the United States Office of Education *Educational Directory*. Monographs and individual pamphlets are indexed as "books" in the *C.B.I.* and carried under the usual listing procedures in the card catalog of the library.

General References

There are a number of good reference books with which the student aspiring to proficiency in the library should be familiar. The time spent in becoming acquainted with these sources is generally well repaid by smoother and speedier prog-

ress in library research. Practice in the use of these basic sources should be included early in the program of the graduate student. The following are particularly comprehensive:

1. *A Guide to Reference Books* (Constance M. Winchell, ed., Chicago: American Library Association, 1960 [formerly, Isadore G. Mudge, ed.]).
2. *Basic Reference Sources* (Louis Shores, Chicago: American Library Association, 1954).
3. *Reference Books: A Brief Guide for Students and Other Users of the Library* (Mary N. Barton, compiler. 4th ed.; Baltimore: Enoch Pratt Library, 1962).
4. *A Guide to References* (Albert J. Walford, London: London Library Association, 1954). Contains many excellent suggestions for the effective use of the library.
5. *How to Locate Educational Data and Information* (*op. cit.*). Contains many excellent suggestions for the effective use of the library.
6. *The Modern Researcher* (Jacques Barzun and Henry F. Graff, New York: Harcourt, Brace, 1957).
7. *How and Where to Look it Up: A Guide to Standard Sources of Information* (Robert W. Murphey, New York: McGraw-Hill, 1958).

The more general references that can be found in any library include:

1. Encyclopedias: *Britannica; Americana; Collier's; World Book* (Juvenile); *Columbia* (Encyclopedia in One Volume); and the *Lincoln Library of Essential Information* (1950).
2. Dictionaries: Funk and Wagnall; *Webster's New Third International; Oxford English Dictionary; Roget's Thesaurus of Words and Phrases;* and *Rodale, the Word Finder.*
3. Almanacs, such as *World Almanac* (New York: New York World, 1868–1931; New York: World Telegram, 1932–date); *Information Please Almanac* (John Kieran, ed., publisher varies, 1933–date); and *Statistical Abstracts of the United States* (Washington, D.C.: Department of Commerce, 1897–date).

A number of similar references pertain more specifically to the field of education.

1. Encyclopedias: the best known and most useful of which is undoubtedly the *Encyclopedia of Educational Research* (Chester W. Harris, ed., 3rd. ed.; New York: Macmillan, 1960

[Walter S. Monroe, ed., 1940 and 1950]). Issued at ten-year intervals, it constitutes the primary tool in the hands of the educational researcher. The *Review of Educational Research* (Washington, D.C.: American Educational Research Association, 1931–date) acts as a supplement to keep it current within a three-year span. The *Education Index* and the various periodicals can be used to fill in the gaps between issues of the *Review* and to deal with topics not covered therein.

Other important encyclopedias include: *Educator's Encyclopedia* (Englewood Cliffs: Prentice-Hall, 1961); *Encyclopedia of Modern Education* (Henry D. Rivlin and H. Schueller, eds., New York: Philosophical Library, 1943); *Cyclopedia of Education* (Paul Monroe, ed., New York: Macmillan, 1911–13, 5 vols.); *Encyclopedia of Child Guidance* (Ralph B. Winn, ed., New York: Philosophical Library, 1943); *Encyclopedia of Vocational Guidance* (Oscar J. Kaplan, ed., New York: Philosophical Library, 1948, 2 vols.); and *Encyclopedia of the Social Sciences* (Edwin R. A. Seligman, ed.). *The Annual Review of Psychology* (Paul Farnsworth, ed., Palo Alto: Stanford University, 1950–date. [Calvin P. Stone, ed. Vols. 1–6]) constitutes the most up-to-date encyclopedia concerning the major aspects of psychology. The 75-volume *Library of Education* (Washington: Center for Applied Research in Education, 1962–5), currently under publication, will undoubtedly constitute a major contribution to the cause of education.

2. Dictionaries: The most adequate is Good's *Dictionary of Education* (New York: McGraw-Hill, 1959) which covers some 17,000 terms in education and related fields. Similar, but somewhat less comprehensive, is the *John Dewey Dictionary of Education* (Ralph D. Winn, ed., New York: Philosophical Library, 1959).

   Other useful dictionaries include: *Comprehensive Dictionary of Psychological and Psychoanalytical Terms* (Horace B. English and Ava C. English, New York: Longmans, Green, 1958); *Dictionary of Sociology* (Henry P. Fairchild, ed., New York: Philosophical Library, 1944); and *A Dictionary of Statistical Terms* (Maurice G. Kendall and William R. Buckland, eds., London: Oliver & Boyd, 1957). Somewhat more specialized, but of fundamental importance to counselors is the *Dictionary of Occupational Titles* (2nd. ed.; Washington, D.C.: United States Employment Service, 1949), which lists some 30,000 job titles and 25,000 job descriptions. Rather similar is

Occupational Literature (Gertrude Forrester, ed., New York: Wilson, 1958), which lists over 4,000 books and pamphlets by occupation.

3. Yearbooks: the *Yearbook of Education* which has both an American (Joseph A. Lauwerys, *et al.*, eds., New York: World Book, 1953–date) and a British (G. B. Jeffery, ed., London: Evans Bros., 1932–40, 1948) version. Each issue of the American edition is devoted to a single topic; the British edition contains a variety of signed articles on all phases of education. *International Yearbook of Education* (Geneva: International Bureau of Education, UNESCO, 1948–date); *Mental Measurement Yearbooks* (Oscar Buros, ed., Fifth Mental Measurements Yearbook, Highland Park: Gryphon Press, 1960); *Statistical Methodology Reviews* (Oscar K. Buros, ed., 2nd. ed.; New York: Wiley, 1951); and the yearbooks of many educational societies ranging from such general societies as the National Society for the Study of Education to the more specialized, such as the American Association of School Administrators and the national associations of teachers of mathematics, English, and so on. An older set is *Educational Yearbook of the International Institute* (Isaac L. Kandel, ed., New York: Bureau of Publications, Columbia University, 1928–44, 20 vols.).
4. *Schoolman's Almanac* (New York: Educator's Washington Dispatch, 1947–date). Not only lists the significant educational events of the previous year and presents the calendar of events for the current year, but also analyzes important educational developments and trends.
5. Bulletins, manuals, and guides: *Requirements for Certification of Teachers, Counselors, Librarians, and Administrators in Elementary and Secondary Schools and Junior Colleges* (Robert C. Woellner, *et al.*, eds., Chicago: University of Chicago Press, 1894–date); *A Manual on Certification Requirements for School Personnel in the United States* (Washington, D.C.: United States Office of Education, 1957); *Education for the Professions* (Lloyd E. Blauch, ed., Washington, D.C.: United States Office of Education, 1955); *Lovejoy's Complete Guide to American Colleges and Universities* (Clarence E. Lovejoy, ed. 6th rev. ed.; New York: Simon & Schuster, 1961); *A Guide to Graduate Study; Programs Leading to the Ph.D. Degree* (Frederick W. Ness, ed., Washington, D.C.: American Council on Education, 1960); *American Universities and Colleges* (Mary Irwin, ed., 8th ed.; Washington, D.C.:

American Council on Education, 1960); *American Junior Colleges* (Jesse P. Bogue, ed., 5th ed.; Washington, D.C.: American Council on Education, 1960); *Atlas of Higher Education in the United States* (John D. Millet, ed., New York: Columbia University, 1952); *Guide to Guidance* (Martha E. Hilton and Ellen P. Fairchild, eds., Syracuse: Syracuse Press, 1960); and *Keys to Professional Information for Teachers* (Roy C. Bryan. Kalamazoo: Western Michigan University, 1957).

Government Documents

The list of government publications is so extensive and so specialized that they generally form a special division of the library organization. Much of this volume of publication is in pamphlet form and is issued under the name of the issuing department and numbered according to series and sub-series. As a result, the location of a specific item by the amateur is frequently a relatively difficult task, though the task is one that can be mastered with a little practice. Such sources as Alexander and Burke (*op. cit.*) give a rather thorough orientation to the field.

All publications of the Federal Government are listed according to the issuing agency in the monthly catalog of the *United States Government Publications.* A subject index is provided monthly and cumulated yearly. Other references are:

1. *Education Index* (*op. cit.*).
2. *United States Government Publications* (Anne M. Boyd and Rae E. Rips, New York: Wilson, 1949).
3. *A Popular Guide to Government Publications* (W. Phillip Leidy, New York: Columbia University Press, 1953).
4. *Subject Guide to United States Government Publications* (Herbert S. Hirshberg and Carl H. Melinat, Chicago: American Library Association, 1947).
5. *United Nations Document Index* (New York: United Nations, 1950–date); and *Educational Abstracts* (UNESCO, *op. cit.*).

Locating state and municipal documents is even more difficult, not so much because of the quantity, but because there is a lack of systematic indexing and of a centralized agency responsible for the cataloging and circularizing of the booklets and pamphlets issued by the various departments. Perhaps the

best sources of such information are the *Book of the States* (Council of State Governments, Chicago: The Council, 1935–date) and the *Municipal Year Book* (Chicago: International City Manager's Association, 1934–date). A number of state publications are listed in the Library of Congress *Monthly Checklist of State Publications* (Washington, D.C.: Government Printing Office, 1910–date). Most studies conducted by the state are filed in the library of the state university, and the librarians there are a fertile source of information on such studies. Some of the state and local studies pertaining to education may be found in the *Education Index,* but there is no systematic provision for this.

Biographical information

Biographical data on important persons are available in a wide variety of sources, most basic among which is any encyclopedia. Sources more specifically "biographical" include:

Who's Who (London: Black, 1849–date) and *Who Was Who* (London: Black, 1916–date); *Who's Who in America* (Chicago: Marquis, 1899–date) and *Who Was Who in America* (Chicago: Marquis, 1897–date); *Dictionary of American Biography* (New York: Scribner's Sons, 1928–36, 20 vols.); *Webster's Biographical Dictionary* (Springfield: Merriam, 1943); *National Cyclopedia of American Biography* (New York: White, 1891–date); *American Men of Science* (Lancaster: Science Press, 1906–date, 1: Physical Sciences, 2: Biological Sciences, and 3: Behaviorial Sciences); and the *National Register of Scientific and Technical Personnel* (Washington, D.C.: American Psychological Association and National Science Foundation, 1940–date). Data on persons currently in the news are found in *Current Biography* (New York: Wilson, 1940–date), which is cumulated yearly into *Current Biography Yearbook; Who's Who in American Current Biographical Reference Service* (Chicago: Marquis, 1940–date); *Biography Index* (New York: Wilson, 1946–date); and the *New York Times Index* (*op. cit.*).

Biographical information concerning important educators would be found in any of the above sources, and in *Who's Who in American Education* (Robert C. Cook, ed., Nashville: Who's Who in American Education, 1928–date); *Leaders in Education: A Biographical Directory* (Jaques Cattell and E. E. Ross, eds., Lancaster: Science Press, 1948–date); *President and Deans of*

American Colleges and Universities (Robert C. Cook, ed., Nashville: Who's Who in American Education, 1935–date); *Trustees and President of American Colleges and Universities* (Robert C. Cook, ed., Nashville: Who's Who in American Education, 1955-date); and *Directory of American Scholars: a Biographical Directory* (Jaques Cattell, ed., Lancaster: Science Press, 1952).

Educators and Educational Agencies

Probably the most up-to-date source of the mailing address of persons and agencies connected with education are the directories issued by the individual agencies. The most comprehensive list of educators, however, is undoubtedly the United States Office of Education *Educational Directory* (*op. cit.*), which lists officers of state, county, and city schools, colleges and universities, and educational associations. Also of interest here are the recurrent publications of some of the major education associations (as found in *Educational Directory,* Part 4):

Adult Education Association (N.E.A.), *Adult Education;* American Council of Learned Societies, *ACLS Newsletter;* American Council on Education, *Educational Record;* American Educational Research Association (N.E.A.), *Review of Educational Research, Newsletter* (also *Encyclopedia of Educational Research*); American Personnel and Guidance Association, *The Personnel and Guidance Journal;* Association for Childhood Education, *Childhood Education;* Association for Supervision and Curriculum Development (N.E.A.) *Educational Leadership;* Department of Elementary School Principals (N.E.A.); *National Elementary Principal, Yearbook;* Kappa Delta Pi, *Educational Forum;* National Commission on Teacher Education and Professional Standards, *Journal of Teacher Education;* National Council of Teachers of English, *Elementary English, English Journal, College English;* Ohio State University (Bureau of Educational Research), *Educational Research Bulletin;* Phi Beta Kappa, *The American Scholar;* Phi Delta Kappa, *Phi Delta Kappan;* and United States Office of Education, *Higher Education,* and *School Life.*

Also in this category are *Patterson's American Education* (Leona H. May, ed., Chicago: Educational Directories, 1904–date), which lists both public and private school officials; *American Universities and Colleges* (*op. cit.*) and *American Junior Colleges* (*op. cit.*); and the directories of various edu-

cational organizations, most of which are indexed in the *Educational Index* under Directories, educational. They can also be located through the card catalog of the library. The best source of the addresses of publishers is the *C.B.I.* and the *Publishers Trade List Annual.* The *Education Index* lists the address of the publishers of the books and pamphlets which it indexes in its paperback issues. The card catalog, of course, lists the city of publication of the volumes cataloged as part of the imprint notation. The address of business firms dealing in educational materials is best found in *School Supply and Equipment Directory* (New York: School Management, 1934–date). At the local level, one should consult the city directory or the yellow pages of the telephone directory.

For addresses and general information concerning educational agencies and institutions (particularly of higher learning), one can consult:

Encyclopedia of American Associations (Detroit: Gale Research, 1956, supplement, 1957); *Directory of Research Agencies and Studies* (Raymond J. Young, Bloomington: Phi Delta Kappa, 1959); *Educational Foundations and Their Fields* (New York: American Foundations Information Service, 1955); *Baird's Manual of American College Fraternities* (George S. Lasher, ed., Menasha: George Banta, 1957); *American Library Directory* (Ann J. Richter, ed., New York: Bowker, 1951); *Patterson's American Education; American Universities and Colleges; American Junior Colleges;* and others previously cited. UNESCO provides a directory of educational organization throughout the world in *Educational Abstracts* (*op. cit.*).

News Items

The primary source of news items is, of course, the newspaper, and all newspapers retain back copies. Magazines also carry news, but some time after its occurrence. For research in this area, it is necessary to use some index that will enable the investigator to locate news items, even after a considerable interval of time has lapsed. Probably the most adequate indexes of newspaper items are the *New York Times Index* and *Facts on File* (1940–date), the latter being best described as a current encyclopedia indexed and cross-referenced semimonthly, monthly, quarterly, and annually. Other sources of a

somewhat more specific nature include the *American Jewish Yearbook* (Morris Fine, ed., American Jewish Committee, Philadelphia: Jewish Publishing Society of America, 1899–date) and the *Negro Handbook* (New York: Current Reference Publications, 1942–date). Of special interest to educators is *Schoolman's Almanac* (*op. cit.*) which contains many news items about educational events. Cartoons in newspapers and magazines can be located in the *Education Index* (under Educational Cartoons) and in the *Cumulative Book Index* (under Caricatures and Cartoons).

Bibliographies

Particularly helpful in the early stages of the review of the literature are the many excellent bibliographies that have been prepared on a number of educational problems. Although many of these are not exhaustive, and, of course, all are in various degrees of out-of-datedness, they can nevertheless save untold hours of searching. In a sense, the *C.B.I.*, the *Education Index*, the card catalog, and many of the other references previously cited constitute bibliographies. Similarly, the extensive lists of references at the end of articles in the *Encyclopedia of Educational Research* and other sources very frequently constitute excellent bibliographies from which to start the research on the literature on a given topic.

Probably the most comprehensive reference to bibliographies is the *Bibliographic Index* (New York: Wilson, 1938–date) which is really a bibliography of bibliographies. It reviews some 1,500 books and periodicals on a semi-annual basis with annual and larger cumulations. Somewhat more specialized is *Guide to Catholic Literature* (Walter Romig, ed., Washington, D.C.: Catholic Library Association, 1888–date). Other bibliographical sources include *Good References* (Washington, D.C.: United States Office of Education, 1931–45); *Bibliographies and Summaries in Education to July 1935* (Walter S. Monroe and Louis Shores. New York: Wilson, 1936); and *American Bibliography* (Charles Evans, ed., Worcester: American Antiquarian Society, 1930–4. Vol. 13; New York: Bowker, 1955). Special bibliographies are also found in such journals as the *Elementary School Journal* and *School Review*. *Historical Bibliographies* (Edith M. Coulter

and Melanie Gerstenfeld, Berkeley: University of California Press, 1935) provides an annotated source of historical references up to 1935. Useful bibliographies on research methods are to be found in the *Journal of Educational Research* (1929–45) and the *Phi Delta Kappan* (1946–date).

Educational Research Studies

Research studies, as such, are not listed separately in most sources. The *Education Index*, for example, is inclusive rather than selective and critical; it lists major research studies and expressions of unverified opinion side by side. Since both may have value depending on the student's needs and purposes, the *Index* brings both to the attention of the reader; it is up to him to check each reference individually for whatever worth it may have for him. Undoubtedly, the *Education Index* is the most comprehensive source of research studies, even though it is not restricted to a listing of such studies alone.

Obviously, the most adequate source of educational research studies is the *Encyclopedia of Educational Research*, which covers the significant research on major educational topics and interprets and provides an extensive bibliography of the research on the topics covered. Complementing the *Encyclopedia* is the *Review of Educational Research*, which reviews the research on a series of fifteen topics in three-year cycles and provides extensive bibliographies of the research literature.

Another source of research studies of special interest to educators—particularly administrators—is the *N.E.A. Research Bulletin* (1923–date) which reports N.E.A. sponsored research in such areas as school enrollment, teacher salaries, teacher supply and demand, and other matters affecting the profession. N.E.A., through its Educational Research Service, also publishes *Questionnaire Studies Completed* (1928–date).

Probably the most comprehensive single source of research studies in education is found in the lists of the theses and dissertations conducted in partial fulfillment of degree requirements. These are becoming progressively easier to locate. The most complete reference to doctoral dissertations is *Dissertation Abstracts* (Ann Arbor: University Microfilms, 1955–date), which abstracts a progressively greater percentage of the doctoral dissertations conducted in degree-granting institutions.

Originally *Microfilm Abstracts* (1938–1955), it took over *Doctoral Dissertations Accepted by American Universities* (New York: Wilson, 1935–55). All dissertations abstracted are available on microfilm and, in many cases, on Zerox at a nominal price. Yearly indexes are provided.

Other important current references to doctoral dissertations include: The *Education Index,* which, under "Dissertations, academic" lists the dissertations abstracted in *Dissertation Abstracts* and the booklets and other series of research studies published by the different schools not making use of the microfilming service of University Microfilms. The latter series generally gives abstracts of the studies completed. Unfortunately, individual titles generally are not indexed except through the table of contents of each volume. An exception to this is found in such notable series as *Teachers College Contributions to Education* (1905–51). Since each dissertation in the series was bound separately, it is listed by author in the *Education Index* and in the card catalog. These series are, of course, becoming scarce, as more and more of the graduate schools of the country have their dissertations microfilmed through University Microfilms. Other sources are:

Research Studies in Education: A Subject Index of Doctoral Dissertation, Reports, and Field Studies, 1941–51 (Mary L. Lyda and Stanley B. Brown, Boulder: The Authors, 1953; Continued, Bloomington: Phi Delta Kappa, 1952–date), and *A Quarter Century of Educational Research in Canada; An Analysis of Dissertations in Education Accepted by Canadian Universities,* 1930–1955 (Willard Brehaut, ed., Toronto: University of Toronto, 1958).

Somewhat older references to doctoral dissertations include:

American Doctoral Dissertations (Washington, D.C.: Library of Congress, 1921–38); *Bibliography of Research Studies in Education* (Washington, D.C.: Government Printing Office, 1926–40 [discontinued during the war]); *Guide to Bibliographies of Theses* (Thomas Palfrey and Henry E. Coleman, Chicago: American Library Association, 1940. With corrections by Rosenberg in *Bulletin of Bibliography,* 18: 181–2, 1945, and 18: 201–3, 1946); *Monroe's Ten Years of Educational Research,* which covers dissertations for the period 1918 through 1927 (Urbana: University

of Illinois Press, 1928); and *Guide to Research Sources in Education* (Emil Greenberg, New York: New York University, 1948).

Master theses are not as completely covered by indexes and are not so readily and systematically referenced. Nevertheless, a number of excellent references exist:

Master Theses in Education (Tom A. Lamke and Herbert M. Silvey, eds., Cedar Falls: Research Publications, 1952–date); *Education Index* under "Dissertations, academic"; and *Master's Theses in Health, Physical Education and Recreation* (Thomas K. Cureton Washington, D.C.: National Education Association, 1930–date).

Occasionally, the author publishes an article in a regular journal outlining the major points of his study. The United States Office of Education has a limited number of theses and dissertations available for loan and, as previously mentioned, some universities maintain summary series of their unpublished master's studies.

Non-degree connected research is much more difficult to find. In fact, only haphazard success can be expected in locating such studies, whether they involve research conducted by university faculty, by military personnel, or by school systems. Some of this research is published in regular journals, but the bulk probably frequently remains relatively unknown and frequently unused. There are, of course, notable exceptions, for example, the publications of such bureaus of educational research as those of Ohio State, Indiana, the City Schools of New York, of Baltimore, to name but a few, whose research is readily available.[2]

Educational Statistics

The most comprehensive source of statistics on all phases of American education, particularly at the national level, is the *Biennial Survey of Education in the United States* (Washington, D.C.: United States Office of Education, 1917–date). This is complemented by special pamphlets and bulletins which are available on request from the Department of Health, Educa-

[2] The California Advisory Council on Educational Research prepares a bulletin listing the studies sponsored by California school districts, county offices, and other educational groups. (See California Teachers Association, *Research Bulletin* No. 153, Burlingame: The Association, April 1962).

tion and Welfare, as well as from other departments and agencies of the federal government. Other useful sources include: *Statistical Abstracts of the United States* (*op. cit.*); *Economic Almanac* (National Industrial Conference Board. New York: Crowell, 1940–date); which gives information of a nature similar to that contained in *Statistical Abstracts;* and the various publications of the Bureau of the Census (see *Catalog and Subject Guide,* 1946–date). In addition to the regular census figures which it provides, the Bureau of the Census issues periodic reports on inter-census estimates of population, data on the labor force, educational levels, income and expenditures, housing, and so on. On the years ending in "2," for example, it gives special statistics on independent school systems.

Probably the best source of statistics on education at the state level is the *Book of the States* (Council of State Government, 1935–date). At the municipal and individual school-system level, little can be done to locate research except to write directly to the local directors of research. Young's *Directory of Educational Research Agencies and Studies* is useful in this connection, both in identifying some of the major studies that have been conducted and in providing the address of the research personnel to whom a request for information may be forwarded.

SUMMARY

1. The concept of science as a series of successive approximations to the truth and the consequent need for the researcher to build on the efforts of previous investigators make it imperative for him to be thoroughly familiar with the writings in the field. A thorough review of the related literature is an integral part of the conduct of research, helping the researcher in the clarification of his problem and the avoidance of duplication, the formulation of insightful hypotheses, the planning of an adequate research design, and the rigorous and insightful interpretation of his findings.

2. The library is a relatively unlimited storehouse of knowledge. Because of the magnitude and complexity of the material housed therein, the student interested in its effective use must become familiar with its organization. Effectiveness in library work calls for speed, accuracy, and dependability in (1) locating the necessary source; (2) deciding what is to be extracted from each source; and (3) taking whatever notes are needed.

3. The review of the literature must be systematic and thorough, or it will produce inadequate results. It is particularly

important for the researcher to have a clear conception of his problem so that he will keep the review of irrelevant material to a minimum while, at the same time, ensuring a complete coverage of what is relevant.

4. It is generally best for the researcher to orient himself to the general nature of his problem through such general sources as a textbook and the *Encyclopedia* (and the *Review*) *of Educational Research* before investigating more isolated references to be traced through indexes.

5. Library holdings fall under two major classifications: *books* and *periodicals*—with a third category of miscellany. The card catalog is the key to the *books* holdings of the library; each volume always has an author card, and generally has a title card and one or more subject cards. Books are generally classified according to the Library of Congress or the Dewey Decimal system, although some libraries have a modified system of their own.

6. Periodicals are generally housed in open shelves in the Periodicals Room. Their effective use is predicated on the use of an index to identify the articles on the subject under study. The most important index for researchers in education is undoubtedly the *Education Index* which lists—by author and subject—the vast bulk of the materials of interest to educators.

7. A wide variety of indexes and general references can be found to cover almost any area in which the modern researcher might be interested. He would do well to develop a certain familiarity with the more pertinent of these sources. General library sources such as Alexander and Burke's *How to Locate Educational Information and Data* should be consulted for special problems.

PROJECTS and QUESTIONS

1. If you have not already done so before, arrange for a guided tour of the library. Visit the stack areas in which books in education and in related disciplines are shelved.
2. a) Identify current leaders in the various areas of specialization, —counseling, curriculum, reading, and so on.

   b) Identify the pioneers in the field of educational research and their contributions. Justify your selection by listing their major contributions (for example: Thorndike: laws of learning, theory of identical components, development of tests and measurements, and so on).
3. Locate the report of a good research study from *Dissertation Abstracts*. Obtain the microfilm and analyze the study from the standpoint of the problem (statement, delimitation, and justification), hypothesis, research design, findings and conclusions, and implications and significance for educational practice.

Whatever contributions statistics can make to the whole problem lies not so much in the provision of cook-books by which problems are solved, but in providing a framework and a way of thinking about problems.

OSCAR KEMPTHORNE

6 Statistical Considerations

This chapter presents an orientation to some of the more basic statistical concepts necessary for the conduct of research. It makes no attempt to deal with their derivation, nor does it make any claim to complete coverage. This is an area in which the serious student of research needs to develop more than superficial skill, for proficiency in statistics is as fundamental to adequacy in research as is proficiency in mathematics to success in physics.

INTERPRETATION OF THE RESULTS OF RESEARCH

Research data become meaningful in the process of being analyzed and interpreted. If research is to be productive, therefore, the plans for analysis must be laid at the time the study is selected and designed, for unless the analysis of the data can be made sufficiently precise to permit interpretations and generalizations, there is no point in conducting the study. The analysis of research data follows rather closely the development of science, some of the principles of which will be repeated in brief for the sake of continuity in discussion. A prerequisite to interpretation is experience, which, bears on the fundamental problem of obtaining accurate and adequate data, for no

conclusion, regardless of the adequacy of analysis, can be more adequate than the data on which it is based. Implied here is the need for a thorough grounding in the area of tests and measurements, statistics, and research, and the principles governing the derivation of adequate data.

The first step in the analysis of research data is categorizing or classifying. Classification is always based on one's purpose and, as it applies to research, should be guided by a hypothesis which provides the framework for the classification. In practice, it is best to classify data into the finest sub-categories one may need, since sub-classes can be combined to give more major classes, but major classes cannot be broken down into finer classes except through retabulation.

Generally, data are most easily processed when they are converted into numerical values. Quantification not only facilitates their manipulation, but also increases the precision with which they can be analyzed. On the other hand, this immediately raises the question as to what to include and on what basis—for instance, "price" must be defined as "wholesale price," "retail price," or other unambiguous notation. Even such elementary aspects as the number of rooms in a house, a person's age, and the number of students in a given university are subject to some degree of misrepresentation arising from a lack of clarity regarding the basis of classification. Before analysis can proceed, it is also necessary to decide whether cases for which complete data are not available should be eliminated or data "manufactured" to replace what is missing. Another problem that might arise is the extent to which one is justified in rejecting apparently incorrect scores or scores that are outside expectation. All of these are rather complicated problems calling for considerable research insight.

Although quantification is a fundamental step in the analysis and interpretation of data, it is not an end in itself, and conclusions must always be interpreted on the basis of the variables being investigated, rather than on the basis of their numerical values. A statistic is an abstraction used to replace a large mass of data—it has no meaning of its own. Furthermore, the use of complex statistics where they are not warranted may impress the unsophisticated, but they are misleading and serve no useful purpose. Certainly, they do not improve the study.

Statistics, by synthesizing data, can facilitate the derivation of conclusions. The process of arriving at decisions and the interpretation of the findings, however, must always remain a matter of logic and judgment, and, for this reason, research must always be directed by a person familiar with the field rather than by a statistician alone. For example, if an increase in the number of spelling errors were to attend a course in creative writing, someone unfamiliar with the fact that creative writing encourages students to write more—and therefore, to make more spelling errors—might be misled into concluding that creative writing is conducive to poor spelling.

The processing of numerical data through statistics calls for competence in the use of statistical methods and for awareness of the assumptions that underlie their development and their application. In order not to mislead or be misled, the researcher must know the strengths and the weaknesses of the statistics which he uses. The investigator must first remember that any statistic is no more accurate than the data on which it is based; statistical manipulation does not endow data with precision which they did not have in the first place. It also must be recognized that, if one decides to ignore assumptions or to use inappropriate measures, he can have his data "prove" anything he wants to prove. Although some people use statistics as a drunk uses a lamppost—for support rather than for enlightenment—this is not a failing of statistics, but rather of the misguided or unscrupulous people who misuse a perfectly legitimate and useful tool.

STATISTICAL CONCEPTS

Statistics as a Tool of Research

Statistics is an indispensable tool for both the consumer and the producer of research; without it, one cannot even read the professional literature intelligently. It does not seem unreasonable, therefore, to expect the holders of an advanced degree in education to possess the statistical competence necessary to conduct simple research into educational problems. Statistics is not particularly difficult when studied systematically, and anyone with an understanding of high-school algebra need not be unduly restricted in understanding statistical con-

cepts, and even in using most statistical procedures in the analysis of research data. On the other hand, it would seem generally advisable to place the emphasis of courses in educational statistics on applications and proper use—with due caution with respect to underlying assumptions and limitations—rather than on mathematical derivation.

Descriptive Statistics

The broad field of statistics can be divided into two major areas: *descriptive statistics* and *statistics of inference*. Descriptive statistics are devised for synthesizing data to describe the status of the phenomenon under consideration. The superintendent, for example, may be interested in knowing that, on the average, each classroom uses 1.2 boxes of chalk per month. Or it may be desirable to synthesize the IQ's obtained by individual students into an overall average for the whole student body. The measures of descriptive statistics most commonly used in education are the *mean*, the *standard deviation*, and the *coefficient of correlation*, each of which can be extended into other phases of statistical reasoning. The mean and the standard deviation, for instance, lead directly into the concept of the *normal probability distribution*, which is particularly important as the vehicle for introducing the concept of statistical significance. The coefficient of correlation has direct bearing on predictive studies and, of course, on factor analysis. Adequate treatment of the nature and purpose of these basic techniques can be found in any introductory text in statistics and nothing needs to be added here. The computational aspects will also be left to other sources. It should, however, be realized that, though computational proficiency is not a prerequisite to the use of statistical procedures as a research tool, a real understanding of these concepts is frequently best promoted through actual practice in computation.

Statistics of Inference

More pertinent from a research point of view are statistics of inference. Research is generally conducted by means of a sample on the basis of which generalizations concerning the population from which the sample was obtained are reached. More specifically, the investigator computes certain sample

values as the basis for inferring what the corresponding population values might be.

Underlying such an extension of a sample value to the corresponding population value is the fundamental concept of *probability*. It is well known that a sample taken at random from a given population will provide sample values that do not agree exactly—except perhaps through coincidence—with those of the population. One must, therefore, make allowances for the operation of chance in any inference relative to a population value based on an obtained sample value. This always involves an element of a risk, and any generalization reached must be made on the basis of probability—never certainty. Furthermore, the investigator must fully realize the possibility—indeed, the likelihood—that he will be in error in a certain percentage of the decisions he bases on statistical inference. Chance fluctuations will invariably cause discrepancies to occur between sample data and expected values; it is the purpose of statistics of inference to help isolate differences that are real from those that are due to chance fluctuations. Generally, the investigator starts by postulating that the results obtained are those occurring through chance effects of uncontrolled variables, a hypothesis which he proceeds to subject to statistical test.

In the interest of clarity, it might be well to identify the following terms as they apply to a sample and a population, respectively, and the symbolism used to represent them.

	Value	*Mean*	*Standard Deviation*	*Number of Cases*
POPULATION	Parameter	μ	σ	N
SAMPLE	Statistic	$\overline{X}$	S	n

This terminology is fairly standard and needs to be understood clearly; thus, the mean of a sample (referred to as $\overline{X}$) is a statistic; the corresponding population mean is a parameter and is represented by the Greek letter μ.

Statistical Probability. The concept of statistical probability is perhaps best presented through what is known as the *binomial distribution.* If ten coins are tossed simultaneously a total of 1024 times, the number of heads in each of these 1024

throws will make a distribution that centers around 5 heads and 5 tails. However, in the 1024 tosses of the ten coins, there would be instances of 6 heads or 6 tails, of 7 heads or 7 tails, and even of 8 heads or 8 tails. In fact, theoretically, there could be as many as 10 heads in a single toss and 10 tails in another. The distribution that one might get on the basis of probability is shown in Figure 6–1.

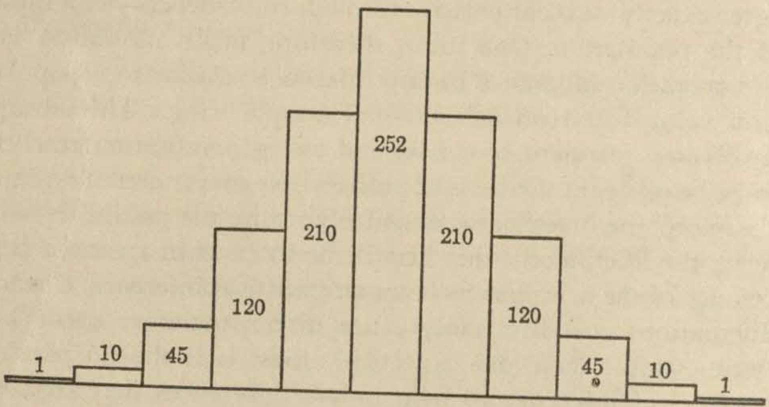


Figure 6–1

With this as background, suppose a person were to toss ten coins only once: certainly, he would get one of the 1024 possibilities, and it could be any one from that showing 10 heads to that showing 10 tails. Supposing that he gets 9 heads—Is this so different from the expected value that he might have reasons to suspect the operation of factors other than chance? Although it is recognized that the operation of chance will occasionally lead to unusual results, unusual results also occur for reasons other than chance, and there are times when the results are so unusual that it may be more logical to suspect that factors other than chance are responsible for their occurrence. The investigator has to make a judgment—and whether he attributes the occurrence to the operation of chance or to other factors, he can expect to be in error in a certain percentage of his decisions. It is a question of the level and the kind of risk he is willing to take, and if, for instance, he were to get ten tails on the first toss of ten coins, he might refuse to accept the operation of chance as the most logical explanation of this unusual (improbable) event.

Sampling Distributions

If a statistic is to serve as the basis for making inferences concerning the population parameter, it is first necessary to ensure that the statistic is an unbiased estimate of this parameter —that is, the sample must be a random sample of the population. For the sample statistic to provide the basis for inference regarding the population, it is also necessary that the general behavior of the statistic in repeated random sampling be known. For example, before we can decide whether a mean IQ of 105 for a given sample is indicative of a true superiority of the population surveyed over the general population, we need to know whether a discrepancy of 5 points from the expected 100 is relatively common in repeated sampling from the general population, or whether this constitutes a most unusual event. More specifically, we need to know the *mean* and the *standard deviation* of the distribution of the sample statistic in repeated sampling. This would, of course, vary with the nature of the statistic in question, which, in turn, would depend on the nature of the problem.

A common research problem is to determine whether a sample statistic can be considered to be within the range of random sampling fluctuations of a given population parameter. For example, we might want to determine how the children of Community X compare with the national norm with respect to intelligence. Or we might want to test the relative difference in gains produced under two different methods of teaching. A number of basic formulas have been devised for dealing with problems of this kind; they are essentially mechanical procedures designed to yield answers which can be interpreted in the light of the problem under investigation.

There are really two problems here: one is computational; the other is logical. The more fundamental, of course, is the logical, and, since it appears that the computational aspects very frequently interfere with the understanding of what one is attempting to do and the rationale underlying such a procedure, it seems more profitable to deal first with the logical considerations.

To make the procedure more understandable, let us take

the relatively familiar distribution of the Revised Stanford-Binet IQ's for the population of American-born Whites. Let us simplify the discussion further by ignoring any question that might be raised regarding standardization procedures and by rounding out decimals to give the following parameters: $\mu = 100$ and $\sigma = 16$. The distribution is generally accepted to be normal, with the cases distributed more or less as shown in Figure 6–2.

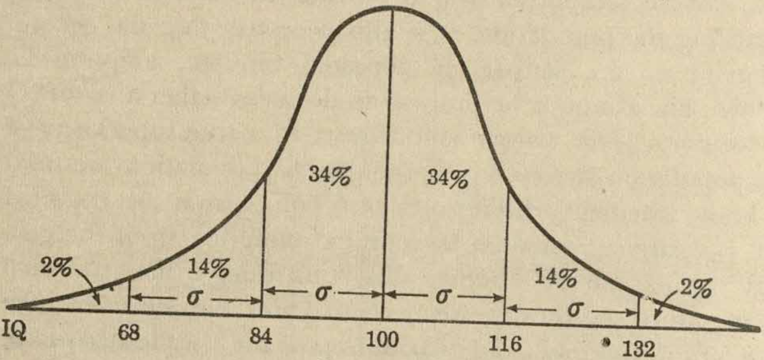


Figure 6–2

Thus, approximately 34 percent of the general population of American-born Whites have IQ's from 84 to 100, and a corresponding 34 percent between 100 and 116. Other percentages can be read directly from the figure, and finer breakdowns can be made by reference to the normal probability table, a copy of which can be found in any textbook in statistics.

If, from this population, repeated random samples of 256 cases are selected completely at random, the IQ of each of these cases obtai d, and a mean IQ computed for each sample, it might be expected that the mean of each of the samples would be fairly close to 100—that is, they would all center around the mean of the population $\mu = 100$, but would depart somewhat from this mean. These sample means would also form a distribution defined by a mean ($\bar{X}_{\bar{x}}$) and a standard deviation ($SE_{\bar{x}}$) of its own. Furthermore, this distribution would approximate the normal distribution, so that it is possible to bracket these sample means within intervals of their standard deviation in exactly the same way as in the distribution of raw scores (Figure 6–2). This point is fundamental to the interpretation of the results of research.

We have seen how the scores are distributed in a normal

probability distribution. For example, 34 percent and 34 percent of the IQ's of the general population fall within one standard deviation (16 IQ points) of 100—that is, between 84 and 100 and between 100 and 116. Nearly 96 percent of the scores fall within two standard deviations of the mean—that is, between 68 and 132. A parallel interpretation can be made with respect to the distribution of sample means in repeated random sampling, except that, instead of the concept of standard deviation used in connection with the distribution of a raw score like the IQ, we must substitute that of the *standard error of the mean*[1] which is defined as $SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ and which, therefore, in the present case has a value of $= \frac{16}{\sqrt{256}} = 1.$

Referring to Figure 6–3, we can establish that 1. the sample means obtained in repeated random sampling from a popu-

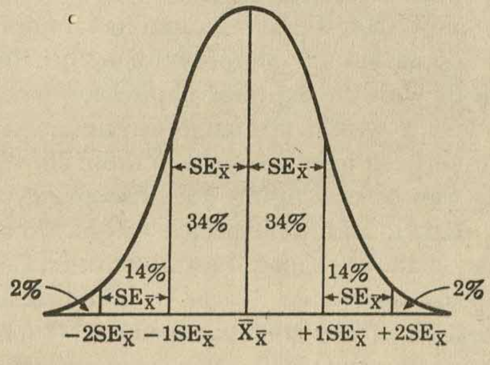


Figure 6–3

lation of mean, $\mu = 100$ and $\sigma = 16$, distribute themselves normally around a mean of their own (that is, $\bar{X}_{\bar{X}}$) very nearly equal to the population mean, μ; 2. this distribution has a standard deviation of 1, so that we might expect 68 percent of our samples to have a mean between $\mu \pm 1\ SE_{\bar{X}}$ (or between $\bar{X}_{\bar{X}} \pm 1\ SE_{\bar{X}}$)—that is, between 99 and 101. We also might expect 96 percent of our samples to have a mean within the range from

[1] The distinction here is that the *standard deviation* pertains to a distribution of raw scores whereas *standard error* is used to refer to the variability of a distribution of sample statistics. The standard error of the mean then is simply the standard deviation of the distribution of sample means in repeated random sampling. The standard error of any statistic is likewise the standard deviation of its sampling distribution.

98 to 102. For a sample to have a mean of less than 98, or more than 102, would be considered a rare event, since this would tend to occur on the basis of chance alone in only two out of one hundred samples in repeated random sampling. A mean of 99.8, for example, would be well within the expected range. Conversely a sample mean of 94 would be expected to occur only rarely through random sampling. Were such a low mean to occur in a single trial, it would be rather difficult to conceive of its occurrence as resulting from chance alone. The exact probability involved could, of course, be read from a table of the normal probability distribution.

The example based on the normal probability distribution is the simplest case, but similar reasoning and corresponding formulas can be used to deal with other statistics. The problem is a matter of simple logic and common sense. We have to explain research results in the simplest and the most convincing manner. We know that whenever random samples are taken from a given population, the sample statistics obtained are not likely to coincide with the expected population parameter, but will tend to deviate within the range of random sampling errors as computed from formulas such as those above. Thus the most obvious way of explaining any discrepancy between a given sample statistic and the expected value is to assume that the discrepancy is due to chance. This assumption can be tested, and, if it appears reasonable, it can be accepted as the most logical explanation of the discrepancy noted. If, on the other hand, the discrepancy noted is so large—for example, a sample mean of 94 in the example above—that it represents an event which is most improbable as the outcome of chance, the investigator must look for a more logical explanation of the difference noted. Thus he might consider this a real difference rather than one that arose through chance, and might conclude that the children of Community X, as a population, are significantly below average in IQ.

THE NULL HYPOTHESIS

Rationale underlying the Null Hypothesis

The investigator always starts out with what is known as the *null hypothesis*—that is, with the assumption that the dif-

ference between the obtained and the expected value is the result of chance. Thus, in keeping with the principle of parsimony which states that phenomena should be explained on the basis of the simplest explanation consistent with all of the facts of the case, the investigator begins with the assumption that the difference noted is due to chance. Specifically, the null hypothesis denies the existence of any real difference between the expected value and that obtained in the sample, until the factor of chance has been eliminated as the causative agent in the discrepancy noted. In an experiment comparing Teaching Method A and Teaching Method B, for example, it is unlikely that the experimental and control groups will make identical gains. By chance alone, one would be likely to exceed the other by at least a small margin. According to the null hypothesis, only when the difference noted in the performance of the two groups is greater than might be accounted for on the basis of chance fluctuations can the investigator assume one method to be superior to the other. On this assumption then, the investigator proceeds to test the difference obtained by calculating the probability of obtaining similar results in repeated random sampling from the same or an equivalent population, where the differences should then be zero. He can reject the hypothesis if the probability of obtaining such a difference on the basis of chance alone is very small—or, of course, accept the hypothesis if the difference is within the range of differences adequately accounted for by chance.

The logic underlying such a test revolves around the probability or improbability of the occurrence (through chance) of a difference of the magnitude of that obtained. If the difference is so large that it makes such an event very improbable, the null hypothesis is rejected, with the implication that a more plausible explanation is to be sought. If, on the other hand, the difference is sufficiently small that its occurrence on the basis of chance is relatively probable, the null hypothesis is accepted, with the understanding that chance *could* account for such a difference. Note that it is simply said tnat chance is an acceptable explanation of the difference obtained. The null hypothesis is never proven or disproven; it is simply accepted as plausible or rejected as implausible.

Similar logic could be used in the comparison of two sam-

ples. For example, in comparing the relative effectiveness of Teaching Methods A and B, if the group using Method A decisively outperforms the group using Method B, it can be concluded that Method A is more effective in promoting pupil growth than is Method B. The rejection of the null hypothesis in this case is synonymous with assuming the possible superiority of one method over the other. Note that what is being tested is not whether there is a difference between the two samples—since this is obvious from the data—but rather the likelihood of a real difference in the *populations* which the two samples represent—that is, in the methods under comparison.

Confidence Levels

The level of improbability necessary to lead to the rejection of the null hypothesis is obviously a matter of judgment, based on the nature of the problem and the risk the investigator is willing to take. Two types of errors are involved here: Type 1 or *Alpha* errors—refer to the acceptance of the null hypothesis when it is actually false. For instance, even though boys, as a population, may be taller than girls at age 15, this may not appear in a sample of 10 boys and 10 girls to a sufficient degree to cause the rejection of the null hypothesis. Type 2 or *Beta* errors, on the other hand, refer to the rejection of the null hypothesis when it is actually true. This occurs when, even though the two populations are actually equivalent—for example, in a study of the standing height of boys and girls at age 11—yet one of the samples turns out to be distinctly superior to the other.

Type 2 errors can be minimized by the simple expedient of rejecting the null hypothesis only when the differences are so fantastically great that the occurrence borders on the impossible, rather than on the relatively improbable. This automatically increases correspondingly the likelihood of the occurrence of Type 1 errors, however, for many sizable sampling differences would then be accepted as being within the realm of chance while, in reality, they reflect real differences in the population under test. Actually, the only two ways in which both types of errors can be reduced simultaneously—and not at the expense of one another—would be by taking larger samples and/or by reducing the sampling variability by selecting a

more restricted population or relying on a matched-pair or improved sampling design.

It is, therefore, a matter of compromise. One type of error must be balanced against the other, and the level of acceptance and rejection of the null hypothesis must be set at what might be considered the most opportune point, depending on the relative severity of the two types of error. Custom and tradition in the fields of education and psychology favor balancing the two types of errors around the points at which there are either 5 chances out of 100 or 1 chance out of 100 of being in error. On a normal probability distribution, for example,

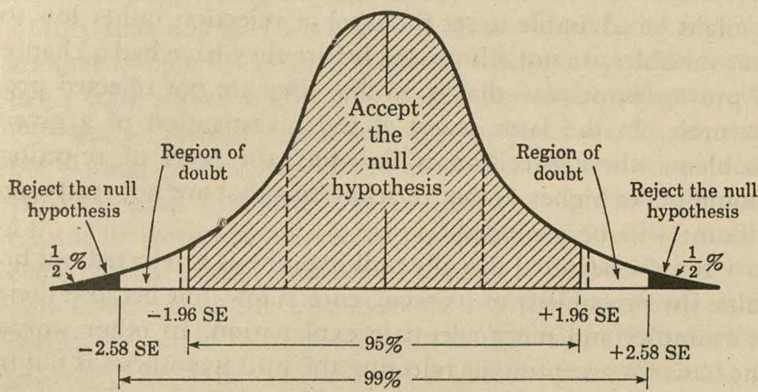


Figure 6–4

critical cut-off points are established at 1.96 standard errors and 2.58 standard errors on either side of the mean to include the 5 percent and the 1 percent level of probability, respectively. Thus, if the difference being tested is such that as large a difference could be expected on the basis of chance 5 or more times out of 100 in repeated random sampling, the null hypothesis is accepted. The difference is satisfactorily accounted for on the basis of chance. If, on the other hand, the difference is sufficiently large that such a difference would occur by chance less than once out of 100 trials, the null hypothesis is rejected on the premise that factors other than chance are probably involved in producing such a large difference. A difference of this magnitude is said to be statistically significant at the 1 percent level—or at the 99 percent level of confidence—which is the equivalent to saying that the probability of a difference of this magnitude occurring through chance is only 1 out of 100, or

less. Differences falling between the 5 percent and the 1 percent level are said to be in the *region of doubt*—sometimes expressed as *significant at the 5 percent level.*

On the normal probability distribution, the critical values that are generally considered are those at 1.96 and 2.58 standard errors on either side of the population parameter, which correspond to the 5 percent and the 1 percent levels of probability, respectively. However, it must be recognized that these are simply arbitrary values, and that other cut-off points could have been selected. In the early stages of exploration, for example, it might be advisable to set the level of rejection rather low so that variables are not eliminated before they have had a chance to prove themselves—that is, so that they are not rejected prematurely. In the later stages of the investigation of a given problem, where precision is essential, the level of rejection should be set higher so that relationships that are not really significant will be excluded.

It must be noted that the null hypothesis is rejected not because the probability of its occurrence is low, but because there is a simpler and more adequate explanation. In other words, the basis for accepting or rejecting the null hypothesis is not so much probability as it is reasonableness—or parsimony. It should be noted further that probability applies not to the null hypothesis—which is either true or false—but to the risk the investigator is willing to take as to the truth or the falseness of the hypothesis.

Tests of Significance

Thus far the discussion has centered around the rationale underlying the testing of the null hypothesis. We now turn to the computational aspects. The formulas for testing the tenability or the unacceptability of the null hypothesis vary somewhat with the type of problem, and with the statistic which is appropriate in each case. Probably the simplest formula concerns the comparison of a sample with a population where the parameter of the latter is known. For example, let us say that the mean IQ of the sample of 256 children from Community X (see page 147) was found to be 101.4 in contrast to the parameter, $\mu = 100$ and $\sigma = 16$, for the general population.

The test of the null hypothesis here is the *critical ratio test,* as follows:

$$\text{C.R.} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{101.4 - 100}{16/\sqrt{256}}$$
$$= 1.4$$

The value of 1.4 in this case is not significant, since it is less than the 1.96 required for significance at the 5 percent level. The null hypothesis is, therefore, accepted—that is, we conclude that the difference, while favoring the children of Community X, is not large enough to exclude at a reasonable level the possibility that this may have been a chance superiority and that another sample of children from the community might just as easily have shown them to be below national average in IQ.[2]

Not all tests of significance can be based on the theory of the normal probability distribution. Certain statistics do not yield a normal distribution in continued random sampling[3] and cannot be interpreted according to the normal model. Statisticians have devised a number of models to deal with the basic non-normal sampling distributions and have prepared tables listing the values corresponding to critical probability levels—generally, the 5 percent and the 1 percent—to permit the acceptance or rejection of the null hypothesis. These table values would not coincide (numerically) with those of the normal distribution, of course, but the rationale underlying their use is the same.

[2] A common variant of the critical ratio test above is the *one-tailed test* which differs from the *two-tailed test* above in that it involves a preconception as to the direction of the difference under test. Thus, in the two-tailed test above, the point was simply whether or not a difference existed between the two means; the direction of this difference was not specified. If, on the other hand, we undertook to test whether, in view of the positive correlation among abilities, we might expect to find tall men to be superior to the general population in IQ, we are interested only in whether they are *superior* in IQ, not in whether they are *different.* The test involved here is identically the same as before, but the critical-ratio values corresponding to a given probability level have to be changed accordingly. Thus, the 1 percent level of significance corresponds to a critical-ratio value of 2.58 in a two-tailed test (that is half of 1 percent on each tail), but 2.33 in a one-tailed test (that is, 1 percent on whichever tail is specified in the hypothesis).

[3] Actually, according to the *Central Limit Theorem,* even when the distribution of a given variable is not normal, the sampling distribution of its statistics, when based on large samples selected at random from a given population, tends to approximate the normal distribution.

Among the more common tests of the null hypothesis, in addition to the normal theory test, are 1. The *t test,* which very closely parallels the critical-ratio test above, except that the standard deviation of the population is unknown, and that the standard deviation of the sample has to be used as a substitute and the results interpreted on the probability table of the *t* distribution. 2. The *F test,* which is used in analysis of variance as a multiple *t* test permitting, for instance, the simultaneous comparison of the gains of three or more groups. It also is used in analysis of covariance where single or multiple experimental and control groups can be compared. It has the added feature of permitting the statistical equating of one or two independent variables in groups in which complete equivalence was not established in the experimental design. 3. The *Chi-square* test which is used when the variables under discussion are in the form of a dichotomy or mutually-exclusive classifications. The Chi-square test appraises the disproportionality among the frequencies of the various cells. 4. *Non-parametric tests* which permit a statistical test of the null hypothesis when the basic assumptions underlying the more rigorous standard tests of significance are not fulfilled.

MODERN DATA-PROCESSING EQUIPMENT

Making use of modern data-processing equipment, especially the electronic computer, in connection with the mechanical and computational aspects of a research study is becoming a standard procedure. This is perhaps most evident in the numerical studies based on multiple regression, analysis of variance and covariance, and factorial designs, where long hours of computation can be reduced to a few hours of careful planning, with a program to tell the machine what to do. Computers are also used in areas where their potentialities are not so obvious. In a documentary-frequency study, for instance, it is now possible to read one theme after another onto a magnetic tape and have a machine produce alphabetical listings of all the words used, along with their frequency. A recent study[4] attempted to determine whether the *Iliad* is the work of a single author, and whether both the *Iliad* and the *Odyssey* were written by the

[4] *Life* Magazine, 52 (August 18, 1961): 41–2.

same author by checking the writings with respect to their similarity in metric pattern.

Probably the most elementary piece of equipment of importance to the research worker in the processing of quantitative data is the desk calculator. Proficiency in the use of a desk calculator is essential for dealing with the smaller tasks connected with the research study, or in planning the more complicated work for the electronic computer. Practice in the use of such machines should be incorporated with advanced training in statistics. The ultimate to date from a research standpoint is, of course, the electronic computer. However, somewhat simpler, and yet of tremendous potential for processing research data of both a qualitative and quantitative nature, especially the latter, is the ordinary punched-card equipment available on almost all campuses.

The Punched Card

The basic unit of operation of the punched-card system is obviously the punched card, or IBM card as it is frequently called, onto which the data are punched. The IBM card has eighty columns, each with twelve punching positions. Those from 0 through 9 are standard digit-punch positions into which raw or coded numerical data can be recorded. The three positions at the top of the card, Positions 12, 11, and 0 (the latter being also a digit-punch position), are zone-punch positions, and are used for alphabetical punches as well as for special signalling—for example, for activating the accounting machine to provide a subtotal for the cards of a given classification.

The first step in the use of the punched card is to code the data so that they can be assigned for punching in certain columns of the card. For example, in the student's master card, the first twenty columns might be reserved for entering the student's name and the next five columns for student number. The twenty-sixth column might indicate sex, while the twenty-seventh might list academic status. On the other hand, if there is any need to conserve column space, Column 26 might be coded: 0–freshman boys; 1–freshmen girls; 2–sophomore boys, and so on. In the next two columns, the last two digits of the student's year of birth might be entered. The remaining columns might be used to record entrance-test data, or any

other information considered useful. Once punched, this information becomes a matter of permanent record. It can be read out of the card as it passes under electric brushes, which activate a current whenever a hole in the card permits the completion of the circuit between the brush and the contact roller on which the card rides.

Numerical data are recorded by a single punch in the regular 0 through 9 digit-punch positions of the proper columns. Alphabetical symbols, on the other hand, require a double punch for each letter; thus, the letter "a" is represented by a dual 12 and 1 punch; the letter "b," by a dual 12 and 2 punch, and so on. The letters "j" through "r" are punched with an 11-zone punch combined with a second punch in positions 0 through 9. The remaining letters of the alphabet are punched with a 0-zone punch and a digit punch in positions 2 through 9.

The Basic Punched-Card Machines

A modern data-processing installation might incorporate a relatively large number of units, including perhaps two or more of the pieces in greatest use. The more basic machines are:

1. The *key punch*, which is operated much like a typewriter, actually punches holes in the proper position of the column assigned to the data. The cards are fed automatically and are aligned in much the same way as margin sets and tabulation stops permit the alignment of data on a typewriter. The card punch has both a reading and a punching station. After a card is punched at the punching station, it moves to the reading station where it stays while the succeeding card is being punched. While the card is at the reading station, it is possible to have any part of the information already punched into it duplicated on the next card by the simple press of the *duplicate key*. Similarly, when a common core of information is to be punched into a number of cards, it is possible to have this information punched automatically by means of a program unit, which will duplicate any part of the information on the program card into any predetermined position on the cards as they enter the punching station. Information specific to each of the cards can be added by individual punching. Since punching can be a tedious task, it is necessary to verify its accuracy. This is done by passing the deck of cards through

a second time, using the *verifier* feature of the card punch, which, instead of punching, simply checks if the holes made in the first run are correctly placed. Any discrepancy between the two operations immediately causes the machine to lock.

2. The *reproducer* permits the reproduction of some or all of the information on a given deck of cards onto another deck, or from a master card to any number of detail cards. The reproducer has both a reading and a punching unit, connected by a control panel, which permits the punches in certain or all of the columns of the master cards to activate similar punches in the same or in different columns of the cards to be reproduced. The reproducer also has a comparing unit which makes possible the comparison of two decks with respect to whatever columns are being checked. Any discrepancy immediately stops the machine.
3. The *sorter* performs the important function of sequencing the cards according to any system of classification into which the data on the cards can be ordered. It can group all the cards with a given punch—for example, it can separate boys and girls, or freshmen, sophomores, juniors, and seniors, and so on. It can arrange all or any of the sub-classes in alphabetical or numerical sequence, as desired, and can select cards out of sequence or cards with a higher sequence number than any specific number. Sorters operate at speeds of from 600 to 2000 cards per minute.
4. The *collator* permits the merging of cards pertaining to the same classification, card by card—for example, it permits the collating of the grades of each student with his master card on the basis of his student number punched on both sets of cards. This could be done by repeated passes of the cards through the sorter, but it would be time-consuming. The collator not only saves time, but also permits the merging of cards on the basis of common information, even though the information is not recorded in the same columns. The machine automatically stops in the event a card is missing or is out of sequence in either one of the decks. The collator also permits the selection of cards having certain punches—for example, the cards of students on probation.
5. The *accounting machine* summates the data in any column or columns and prints sub-totals and totals as ordered. For example, it can list a student's courses and grades, from separate cards, adding his credits attempted, his quality points

earned, by semester, and to date; or it can provide the same information (or simply summary data) for any classification, as for example, for freshmen. When connected to the *reproducer*, it can also have punched summary cards for any classification while, at the same time, printing the data on sheets.

6. The *interpreter* translates the holes in a card into alphabetical or numerical symbols, which it prints on the same card in order to allow for its easy visual identification.

The Electronic Computer

The potentialities of the electronic computer are relatively unlimited. All that is necessary to harness its talents is for the investigator to give it elaborate instructions in the form of a program for each of the many steps that it must take. The machine is nothing but a high-speed idiot which can go through steps with the speed of light—with almost complete accuracy.[5]

Electronic computers fall into two major classes, *analog* and *digital;* the latter is generally the more versatile and adequate for research purposes. The crucial task is the planning of the program, which is made effective as far as the machine is concerned by "reading" the program into its memory unit, and it is now possible to get canned programs for almost any standard statistical procedure, which the investigator can adapt rather readily to his particular set of data.

Electronic computers have made the processing of data by hand or desk calculator essentially obsolete. This is especially true when a large number of cases, or a large number of variables, are to be processed. This is speed and accuracy that cannot be matched by human hands, and every doctoral student working on a project which lends itself to the punched-card or the magnetic tape should certainly avail himself of such facilities. Although the cost per hour seems great, the speed at

[5] The machine is almost completely accurate, even to the point of rejecting its own solution in case of error and, of course, of stopping when it is given incomplete or contradictory orders. The errors that are likely to creep in are those in coding and in punching, both of which tend to be tedious tasks and subject to human error. Generally these processes should be checked by having a second operator go through the various steps independently of the first. It is also essential that the code be made so clear that ambiguity of judgment in coding is reduced to a minimum. Difficulties in coding should be anticipated at the time the study is planned or the pilot study is conducted.

which they operate brings the cost for an overall project below that of clerical help.[6]

Another factor in the use of the computer is that, once the data are placed on cards, any number of cross-comparisons that would not otherwise be made can be run for a modest cost. Whereas, when such comparisons are done by hand, only a fraction of the information inherent in the data is extracted, data-processing machines permit the extraction of every ounce of information of which the data are capable. This can result in uncovering significant relationships which were completely unanticipated.[7]

Students should become familiar with the electronic computer, at least to the point of knowing what it can do, and how to plan their studies and the ordering of their data to capitalize on its potentialities. The modern research worker cannot afford to overlook the possibilities of this fantastic research ally, which can provide in minutes the solution to problems which, a few years ago, would have taken a lifetime to attain.

The availability of such facilities, however, with its possibilities of roboting research puts a new light on research, particularly as it applies to the fulfillment of the degree requirements. What of the student who has the IBM extract preadmission and grade data on freshmen and has the computer derive an equation predicting likely academic success. Does this constitute adequate fulfillment of the research requirements for the degree? There is, of course, no absolute answer. It would seem logical, however, for graduate schools to want their students to do just a little more than can be done clerically. While

6 Because of special educational rental rates given when the equipment is used part-time for unsponsored research, its use is frequently free to students and faculty members working on a dissertation or project.

7 This is not to say that the student should throw everything including the kitchen sink into his dissertation. The rule of being guided by one's problem and one's hypothesis still holds, for, if the student tests every single possibility, one is bound to be significant—simply on the basis of chance. Selecting one's hypotheses after the data have shown them to be significant is an *ex-post-facto* approach at its worst. It is only when they become meaningful with respect to the conceptual framework of the study that these tests can be accepted. On the other hand, such trial-and-error approaches might provide insights that could be used in future studies. Whenever such comparisons are meaningful for the issue of the study, they should be discussed, perhaps in an appendix, and mentioned as a suggestion for further study.

modern aids can relieve the student of drudgery, they should free him to make a greater contribution in the realm of planning and originality so that, rather than doing as much as before with less labor, he ought to be expected to produce more with the same outlay of time and effort.

SUMMARY

1. Statistical proficiency is fundamental to the proper analysis of research data, particularly those of the more advanced stages of the investigation of a complex phenomenon.

2. Descriptive statistics attempt to synthesize data in order to describe the status of phenomena. Statistics of inference is concerned with projecting sample data to provide a judgment concerning the phenomenon as it actually exists.

3. Research deals with a sample from which it derives certain statistics, which it then uses as the basis for inference concerning the corresponding population parameters. Basic to such inferences are the concepts of the sampling distribution of the statistics in repeated random sampling, of probability, and of fiducial limits within which the population parameter can be expected at a given level of confidence.

4. In keeping with the principle of parsimony, the researcher refuses to attribute the occurrence of the phenomenon in question to the operation of the variable under study until the possibility of its having occurred through the operation of chance has been excluded at a given probability level. This is the essence of the null hypothesis.

5. Like all hypotheses, the null hypothesis is never proved; it is simply accepted as plausible (perhaps as one of many plausible hypotheses that could be considered), or rejected as improbable. In this choice, two types of errors are possible: (1) accepting the null hypothesis when it is false; and (2) rejecting the null hypothesis when it is true. Since the risk of one and the other of the two errors varies inversely, it is a matter of balancing one risk against the other. In educational and psychological research, the critical probability levels for the acceptance and rejection of the null hypothesis are generally (arbitrarily) set at the 5 percent and the 1 percent level, respectively. The risk of both types of errors can be reduced simultaneously by increasing the sample size and/or the precision of the sampling design.

6. A difference large enough to cause the rejection of the null hypothesis is said to be significant—that is, not that chance cannot account for such a large difference, but that there is a more parsimonious explanation in the present instance.

7. Modern data-processing equipment is a boon to the modern researcher, permitting the analysis of data—and therefore the in-

vestigation of problems—which a few years ago would have been out of the question. The basic unit of operation of the punched-card system is, of course, the IBM card onto which the data can be punched and which then can be processed quickly and accurately. The electronic computer is even more fantastic in its speed and accuracy and in the complexity of the problems which it can solve. All that is needed to harness its tremendous potential is a program of instruction to tell it what to do and in what sequence.

PROJECTS and QUESTIONS

1. Trace the use of statistics as a tool of educational research. Who are some of the important contributors to its development as a tool of science?
2. Alonzo Grace (A.E.R.A., *Annual Report,* 1962) recommends that college of education faculties be retrained in research and statistical methods. How might this recommendation be effected?
3. Get acquainted with modern data-processing equipment. Obtain information from an expert on the potentialities of the electronic computer for educational research purposes.
4. How might the facilities of the computer be made available to the teachers in the solution of educational problems? (Consider the role of the research bureau in the central office in this connection.)

SELECTED REFERENCES

ALSOP, JOSEPH, *et al. Bibliography on the Use of the IBM Machines in Science, Statistics, and Education.* New York: International Business Machines, 1956.

ANDREE, RICHARD V. *Programming the IBM 650 Magnetic Drum Computer and Data-Processing Machine.* New York: Holt, Rinehart & Winston, 1958.

BAIER, DONALD E., *et al.* "Can Personnel Researchers Test and Train Themselves in Statistics?" *Educational and Psychological Measurement,* 12 (Summer 1952) : 267–74.

COLLIER, RAYMOND O. and STUNKARD, CLAYTON L. "An Analysis of Variance of Multiple Measurements on Subjects Classified in Unequal Groups of One Dimension," *Journal of Experimental Education,* 25 (March 1957) : 255–62.

ENGELHART, MAX D. "The Analysis of Variance and Covariance Technique in Relation to the Conventional Formulas for the Standard Error of a Difference," *Psychometrika,* 6 (August 1941) : 221–33.

FATTU, NICHOLAS A. "Processing of Data," in Harris, Chester W. (ed.) . *Encyclopedia of Educational Research.* 3rd ed.; New York: Macmillan, 1960.

Fisher, Ronald A. *Statistical Methods and Scientific Inference.* New York: Hafner, 1956.

Gourlay, Neil. "F-Test Bias in Experimental Designs in Educational Research," *Psychometrika,* 20 (September 1955): 227–48.

———. "Covariance Analysis and its Application in Psychological Research," *British Journal of Statistical Psychology,* 6 (May 1953): 25–34.

Grant, D. A., *et al.* "Statistical Theory and Research Design," *Annual Reviews of Psychology,* yearly, Vols. 1 (1950) –date.

Harris, Chester W. "Statistical Methods," in Harris, Chester W. (ed.). *Encyclopedia of Educational Research.* 3rd ed.; New York: Macmillan, 1960.

Hotelling, Harold. "The Impact of R. A. Fisher on Statistics," *Journal of the American Statistical Association,* 46 (March 1951): 35–46.

Jackson, Robert W. B. "Some Difficulties in the Application of Analysis of Variance Method to Educational Problems," *Journal of Educational Psychology,* 32 (September 1941): 414–22.

Johnson, Palmer O. "Modern Statistical Science and its Function in Educational and Psychological Research," *Scientific Monthly,* 72 (June 1951): 385–96.

———. "Some Promising Newer Methods," *Phi Delta Kappan,* 35 (October 1953): 35–40.

Johnson, Palmer O. and Fay, Leo C. "The Johnson-Neyman Technique, its Theory and Application," *Psychometrika,* 15 (December 1950): 349–67.

Johnson, Palmer O. and Neyman, Jerzy. "Tests of Certain Linear Hypotheses and their Application to Some Educational Problems," *Statistical Research Memoirs,* 1 (1936): 57–93.

Kaiser, Henry F. "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement,* 20 (Spring 1960): 141–51.

Kendall, Maurice G. and Buckland, William R. *A Dictionary of Statistical Terms.* London: Oliver & Boyd, 1957.

Kish, Leslie. "Some Statistical Problems in Research Design," *American Sociological Review,* 24 (June 1959): 328–38.

Kogan, Leonard S. "Variance Designs in Psychological Research," *Psychological Bulletin,* 50 (January 1953): 1–40.

Lindquist, Everett F. "The Iowa Electronic Test Processing Equipment," in *1953 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1954.

Loevinger, Jane, *et al.* "Maximizing the Discriminating Power of a Multiple Score Test," *Psychometrika,* 18 (1953): 309–17.

McCarthy, Philip J. *Introduction to Statistical Reasoning.* New York: McGraw-Hill, 1957.

Ostle, Bernard. *Statistics in Research: Basic Concepts and Tech-*

niques for Research Workers. Ames: Iowa State College Press, 1954.

REMPEL, PETER P. "The Use of Multivariate Statistical Analysis of the Minnesota Multiphasic Personality Inventory Scores in the Classification of Delinquent and Nondelinquent High School Boys," *Journal of Consulting Psychology,* 22 (February 1958) : 17–23.

SAVAGE, I. RICHARD. "Non-Parametric Statistics," *Journal of the American Statistical Association,* 52 (September 1957) : 331–44.

SHEN, EUGENE. "Experimental Design and Statistical Treatment in Educational Research," *Journal of Experimental Education,* 8 (March 1940) : 346–53.

SMALLWOOD, HUGH M. "Statistical Methods in Research," *Journal of Chemical Education,* 23 (July 1946) : 352–6.

SWINEFORD, FRANCES. "Selected References on Statistics, the Theory of Test Construction, and Factor Analysis," *School Review,* 60 (November 1952) : 491–7; 61 (November 1953) : 491–9; 62 (November 1954) : 488–96.

WALKER, HELEN M. "Statistical Inference," in Monroe, Walter S. (ed.) . *Encyclopedia of Educational Research.* New York: Macmillan, 1950.

———. "Statistics: Sense, and Nonsense," *Teachers College Record,* 56 (November 1954) : 68–73.

WRIGLEY, CHARLES. "Electronic Computers and Psychological Research," *American Psychologist,* 12 (September 1957) : 501–9.

All empirical knowledge is, in a fundamental sense, derived from incomplete or imperfect observation and is, therefore, a sampling of experience.

FREDERICK F. STEPHAN

7 Sampling

Probably no concept is as fundamental to the conduct of research and the interpretation of its results as is sampling.[1] Barring the unusual instance in which a complete census is taken, research is almost invariably conducted by means of a sample, on the basis of which generalizations applicable to the population from which the sample was obtained are reached. Even when a complete census is taken, there is generally some thought of this particular population being a sample of future populations to which the results of the present investigation will apply. Indeed, all research can be considered a sample of the multitude of studies that could be done on a given subject. Rarely is research interested in a sample for its own sake.

The Nature of Sampling

At its more advanced levels, sampling can involve a highly complex set of procedures which requires an understanding not only of sampling techniques but also of the mathematics underlying its use. Precision in sampling is par-

[1] This section could logically have been included in Chapter 6. It is presented here after the section on interpretation has presented the basic statistical concepts necessary for its comprehension.

ticularly important in normative-survey research, where the sample must be selected in complete compliance with the principles of sampling, if it is to have meaning for the population. In experimental studies, on the other hand, such correspondence between sample and population is generally not as important, since the crucial point is the relative equivalence of the two groups being compared, rather than the complete agreement of the samples with the population.

Sampling is both necessary and advantageous. Taking a complete census is generally both costly and difficult; in many cases it is completely impossible. What is not so clearly recognized by the layman, who feels that one takes a sample when he cannot get a complete census, is that sampling frequently results in more adequate data than a complete census. In an interview study, for example, sampling not only saves money but also permits greater care and control to be asserted; it allows for better training and co-ordination among the interviewers; it permits greater depth in interviewing; it allows the interviews to be conducted in a relatively short time so that the distorting effects of the passage of time are minimized; it also permits greater depth in analysis and greater accuracy in processing.

Modern statisticians feel that taking a complete census is frequently a sign of statistical (sampling) incompetence. Not only is sampling more practical than a complete census but, by permitting greater control over every aspect of the selection of the cases, it actually produces more accurate results. As Hansen points out, "If we merely wanted to get national statistics, there would be no reason for taking a census every ten years. This could be done more accurately through sampling procedures, and at a fraction of the cost."[2] Sampling is particularly appropriate to situations in which the phenomenon under study is undergoing rapid changes.

This is not to suggest that sampling is desirable in itself or that the smaller the sample, the better. Admittedly, research should be based on a substantial number of cases, but it must be recognized that the increase in precision obtained through increasing the sample size is frequently more than negated by the errors and other difficulties that accompany a wide survey.

[2] Morris H. Hansen, "More than Noses Will Be Counted," *Business Week*, February 27, 1960, pp. 30–1.

To use an obvious example; if a wine taster had to consume his entire consignment in order to determine its quality, he would certainly defeat any purpose tasting the wine is expected to serve. Furthermore, it is very likely that as he proceeded with the task, errors of judgment would probably increase in direct proportion to the decrease in his sobriety. It might be better to settle for a sample—a small sample!

Obviously, a major reason for sampling is to reduce expense—in time, effort, and money—and the factor of cost must be balanced against the adequacy of the data that are obtained. Some of the problems involved include the question of the specific purpose the sample is to serve, the degree of precision the estimates should have, and the funds available to obtain the desired accuracy. Frequently, the money that can be saved by taking a small sample might be more profitably spent in carrying out a pilot study that would make the design of the study more meaningful. Conversely, complicated sampling designs may so increase the cost of processing and analyzing the data that they negate any saving sampling affords. This is especially true if checks must be introduced to forestall error.

If sample data are to be used as the basis for generalizing to a population, it is essential that the sample be representative of that population. While this is a principle with which everyone agrees, it is also a principle that is incapable of implementation. In the strict sense of the term, a representative sample would be a miniature or replica of the population, at least with respect to the characteristic under investigation, if not in all respects. In order to check the representativeness of the sample, therefore, the corresponding population characteristics would have to be known—in which case there would be no need for a sample. The problem is resolved at the operational level by seeking a sample which is random, rather than necessarily representative—that is, a sample which falls within the range of random sampling errors of being representative with respect to the trait under study. The required population parameter then can be estimated, on the basis of probability, to lie within a band or interval centering around the sample value obtained. More correctly, since the latter is the pivotal point, the sample mean is estimated to be within sampling errors of the population mean.

A crucial point here is to define the population from which the sample is to be taken and to which the conclusions of the study are to apply. One must be very suspicious, for instance, of samples that select themselves, such as in a questionnaire study where only a small percentage of respondents reply, or of samples selected simply on the basis of ready availability. In such instances, it is relatively impossible to establish the specific population to which the results of this kind apply—that is, the population of which such a sample might reasonably be considered representative.

ERRORS IN SAMPLING

Classification of Sampling Errors

Sample data that are not representative can suffer from errors of a *random* and/or a *systematic* nature. These errors can be classified further as errors of *sampling* and /or *measurement,* providing a four-way classification which is shown in Figure 7–1. Cell A refers to the unavoidable errors that occur

	Random	Constant
Sampling	A	B
Measurement	C	D

Figure 7–1. Errors of Sampling

whenever sampling is done. If the investigator has decided on a sample of $n = 100$, and has already selected the first 99, the 100th case, selected at random, may be high, low or average in the trait in question, and will, therefore, cause some shift in the sample statistic. These errors tend to cancel each other to the point that, if n is sizable, the sample statistic will tend to stabilize close to the population parameter. Furthermore, not only can these errors be minimized to any fractional value by increasing n but their magnitude can be estimated.

Cell B refers to errors of bias in sampling—that is, sampling errors which do not cancel out but which lean systematically in one or the other direction of the population value. For example, if one were to sample for income by taking every corner house in the city, he would probably incorporate a bias in his data, since people living in corner houses probably draw above average income for them to be able to afford these somewhat favored locations.

To the extent that systematic errors exist, the data are of limited use as the basis for generalizing to the population. For example, to determine the average IQ of a given school by instructing each teacher to select the first two students who complete their assignment, or two honor students, or two students who volunteer would most likely provide fictitious results as far as the overall school status is concerned.

Cells C and D refer to errors in measurement, rather than to errors in sampling. Measurement errors are, of course, involved in any sampling results, since sampling calls not only for the selection of sample cases but also for the determination of their characteristics. The errors in Cell C are those due to the unreliability of the testing. On any measuring instrument, most students are likely to be mismeasured to some degree. These errors cannot be eliminated completely, but they can be minimized even for a given student by basing his score on an extended and comprehensive measuring program that will permit errors to cancel out. They can be minimized further with respect to the sample statistic by having a sizable sample, which permits the self-cancellation of whatever individual (random measurement) errors still remain.

Cell D concerns another bias—that due to systematic errors of measurement. If, in the testing of a sample of students for IQ, the examiner inadvertently allows an extra three minutes for the test, for instance, there will probably be a systematic tendency for the sample statistic to be higher than it should be. And this would be so regardless of the size of the sample for which the extra time was allowed.

Relative Magnitude of Sampling Errors

From the standpoint of research, the "bad" errors are the systematic errors—both in sampling and in measurement. Not

only can the size of random errors be estimated, but they can be reduced to decimal values by the simple expedient of increasing the size of the sample and the reliability of the tests used. It might even be noted that, while unreliability is serious when dealing with one individual—in guidance, for example—these individual irregularities do not affect the overall sample statistic appreciably but tend to cancel out.

The magnitude of random sampling errors as they affect the sample statistic can best be appreciated by referring to the section on the standard error of the mean, $SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, on page 149. Thus, if the average IQ of repeated samples of $n = 256$, taken from the general population, is calculated, chances are 96 to 4 that a given sample mean will fall within two IQ points of the population mean. The size of the random sampling errors which concern the sample statistic depends on the size of the sample, the variability of the trait under study, and the sampling design used. If greater accuracy is desired, therefore, it can be obtained by increasing either the size of the sample or the homogeniety of the variable under investigation, or by using a more adequate sampling design which will decrease the variability of the sampling distribution of the statistic under study for a given sample size.

Systematic errors, on the other hand, are frequently difficult to detect. One cannot tell by looking at a distribution whether or not the condition of randomness was fulfilled, nor is there any test of the randomness of sample data. Furthermore, the size of systematic errors cannot be estimated since they are outside the scope of statistical theory. To make matters worse, such errors can be large. The effects of selecting the corner house as a basis for sampling for income, or of allowing three extra minutes on a standardized test, is bound to cause sizable errors. Similarly, non-returns in a questionnaire study may incorporate a considerable bias, the extent and even the direction of which is sometimes difficult to estimate.

It might be well at this time to distinguish between what might be called the *parameter* and the *true* value for a population. The parameter is probably best conceived as an extension of the sample statistic. As the sample is increased in size to become a complete census—that is, as n leads to N—the sample

statistic becomes the population parameter. Generally, the larger the sample, the closer the sample statistic approaches the population parameter. Thus, using the IQ as the variable, if the investigator takes a sample of two cases, the mean may be a long way from the population parameter. As he increases his sample to three cases, then four cases, and finally *n* cases, the mean will swing back and forth, gradually stabilizing closer and closer to the parameter to the point that, as the sample comes progressively closer to including everyone in the population, the sample statistic comes progressively closer to the population parameter. If there are systematic errors, however, including more and more of the population in the sample will not correct for such errors, and the sample statistic will stabilize near the parameter but not near the true value—if we define the *true* value as an errorless parameter.[3]

If the results obtained are systematically higher or lower than the corresponding true value, the sample is biased and the discrepancy is called an *error of bias*. This is a phenomenon with which even censuses must cope. The United States census, despite a relatively complete coverage, probably does not get a true value for the age of women, any more than the Bureau of Internal Revenue gets a true picture of the actual income of the Americans who file. In both instances, since we are no longer sampling, the discrepancies are due almost exclusively to systematic errors of measurement. Bias can also stem from systematic errors of sampling. Errors of bias are frequently as large as they are unnecessary, and it does not make sense to increase sample size and cost to reduce random errors to the third decimal place, and leave untouched king-sized constant errors. The one thing that is unquestionably more misleading than a small biased sample is a large sample with an equal bias.

Random errors can be reduced by increasing the sample size. They are, of course, eliminated completely when the sample size is increased to include everyone in the population. This can be seen from the formula for the standard error of the mean, for example, which when stated in full, becomes

[3] Mathematicians would probably prefer to think of the population parameter as the true value. It would then be necessary to recognize that the extension of the sample statistic does not necessarily give the parameter.

$$S.E._{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

It can be seen from the factor at the right that as n reaches N, the standard error becomes zero. Increasing the size of the sample can also eliminate systematic errors of sampling, since, obviously, there can be no errors of sampling—random or systematic—when one no longer samples. This would not, however, eliminate errors of measurement, random or systematic.

In practice, the investigator takes only a small fraction of the total population, and he can, therefore, continue with a bias even when the sample is numerically large. A Republican-Party worker, for example, canvasses ten of his fellow party workers regarding the likely outcome of a coming election. Realizing that his sample is too small, he canvasses another ten of his fellow party workers. Even if he extends his sample to 100 by including the families and the close friends of his fellow workers, he is merely stabilizing his bias, not eliminating it. Only through increasing his sample to the point of exhausting the bias by running out of obviously Republican groups will his sample become more representative.

This situation was illustrated by the *Literary Digest* fiasco of 1936. On the basis of a sample of nearly two and a half million questionnaires returned from over ten million mailed to potential voters, selected largely through its own subscription lists, automobile registrations, and telephone directories, the *Digest* predicted the overwhelming defeat of Roosevelt—only to end with a 20 percent error in their prediction. During the same election, *Fortune,* on the basis of a sample of 4500, was able to approximate the actual results within 1 percent, and also to predict the likely error of the *Digest* poll. Apparently, even with a sample of nearly two and a half million, the *Digest* had not exhausted the bias to the point of including a sufficient representation of the unemployed and the lower socioeconomic groups, who were eagerly waiting for Roosevelt's New Deal.

SAMPLE SIZE

Other things being equal, the larger the sample, the greater the precision and the accuracy of the data it provides. And, contrary to the common belief, the precision of the data is deter-

mined by the size of the sample, rather than by the percentage it is of the population. This can be shown directly by the formula of the standard error of the mean, $S.E._{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ for instance. Except in the case of a large sample taken from a small finite population, the precision of the sample mean is determined by the term n in the denominator—and, of course, σ—rather than by the ratio $\sqrt{\frac{N-n}{N-1}}$. In fact, when the population is large, the term at the right of the equation does relatively little to improve precision, and it is generally omitted from the formula.

The size of the sample which he should take is invariably one of the first questions a graduate student asks his advisor. The exact procedure by which to determine the sample size required varies with the nature of the variable and its sampling distribution, but the basic procedure can be illustrated in connection with the mean of repeated random samples based on the normal probability distribution.

As we have seen (page 153), the chances are 95 to 5 that a sample mean in repeated random sampling will fall within the interval of $\mu \pm 1.96\ SE_{\bar{X}}$. The next question is the degree of accuracy expected: Would the purpose of the study be adequately served if the sampling errors were kept within 2 percent at the 95 percent confidence level—that is, would it be satisfactory if the investigator could be confident at the 95 percent level that the sample mean does not differ from the population parameter by more than 2 percent (or 2 points in the case of the IQ)? If this is acceptable, the investigator can use the formula for the standard error of the mean to provide the required value of n.

$$
\begin{aligned}
1.96\, SE_{\bar{X}} &= 2 \\
1.96 \frac{\sigma}{\sqrt{n}} &= 2 \\
1.96 \frac{(16)}{2} &= \sqrt{n} \\
15.68 &= n \\
246 &= n
\end{aligned}
$$

Thus, he would need a sample of 246 cases in order to meet the conditions of a 2 percent error at the 95 percent confi-

dence level. If he insists on a 1 percent error, and further, if he wants to raise the confidence level to the 99 percent level, he will have to increase his sample size by a considerable margin, as indicated by the following relationship:

$$2.58SE_{\bar{X}} = 1$$

Similar computations will provide an estimate of the sample size necessary for obtaining any degree of precision at any confidence level in whatever statistic is being considered. Thus the question of the size of the sample to be selected is answered on the basis of the precision (and the confidence level) desired. If one is content with any approximation to the population parameter, the sample size can be small; if, on the other hand, a greater degree of precision is required, the sample must be correspondingly greater.[4]

To summarize: the answer to the question of the size of the sample that is required is to be found in the margin of error that can be tolerated in the final estimate of the population parameter. Precision in the estimate of the population parameter requires the application of methods of analysis which will extract maximum information from the data that is obtained. But, at the risk of monotony, it must be repeated that it is a fallacy to expect mere sample size to ensure accuracy, since sample size will not generally eliminate any bias inherent in the sampling or measurement techniques. The latter is the area that needs to be watched carefully, for the errors that may occur there can make any attempt at refinement and precision through increased sample size look relatively misguided.

THE MECHANICS OF SAMPLING

Definition of Population

Sampling procedures involve a number of considerations which must be clearly understood if adequate results are to be obtained. The first problem is to clarify the purposes of the study and then, in the light of these purposes, to define the pop-

[4] Since, in most formulas for the standard error, sample size features as the square root of n in the denominator, a doubling of the precision of a sample statistic generally calls for quadrupling the sample size. The precision can also be increased by restricting the population in order to increase its homogeneity, that is, reduce σ.

ulation which is to be sampled. This definition should be sufficiently clear so that there is no question about the inclusion or exclusion of a specific case, or about the applicability of the conclusions to any given case or group. For example, in a study of juvenile delinquency, it must be clearly stated whether the population consists of those who have at one time or another been delinquent or simply those who have been caught.

Although a population can be relatively unlimited—for example, mankind—research must concern itself with *restricted* populations, such as school children, junior-high-school students in the State of . . . , school children in a given county, or perhaps freshmen attending the University of . . . . Generally, the more homogeneous the population from which one samples, the more precise the results that can be derived. Since sample data can be generalized only to the population from which the sample is obtained, however, it is generally inadvisable to over-restrict the population under investigation. It is particularly important not to make a false definition of the population. The investigator cannot, for example, define the population of a given school as "the children present on the day of observation," if his problem is one of the health of these children, since the presence or absence in school of a given child may be related to the status of his health. Sample size must be related to such questions as the nature of the survey, the instrument to be used, and the means of access to the population, as well as to the particular sampling design. Thus, if the sample is to be contacted by a questionnaire, the sample might be larger than if interviews are to be conducted. The unit of sampling is also important. If a group test is to be administered to an entire class at one time, a larger sample might be taken than if individual tests are required. In all cases, the size of the sample should be in line with the degree of precision which is required.

Basic Principles of Sampling

The most crucial problem in sampling is probably the actual selection of the sample. The theoretical considerations underlying this selection—representativeness, randomness, and so on—must be implemented. The task is probably best approached from the basic principle that every member of the population must have an equal chance of being included in the

sample. This immediately poses a number of complications, the first of which is the relative impossibility of obtaining an adequate listing of any given population. The telephone directory is not an adequate listing of the residents of a given city, since some residents are listed more than once, while others are not listed at all; it is not even an adequate listing of telephone subscribers. It is almost impossible to get an adequate list of the students of a given college—unless defined arbitrarily—when one considers those who registered late, those who dropped out recently, those who are carrying a partial load, those who are registered for non-degree classes, and so on. In fact, when considered critically, nearly every listing—the telephone directory, the city directory, the tax rolls, the voters' list, auto registration—is invariably incomplete, inaccurate, outdated, or otherwise inadequate from the standpoint of almost any sampling purpose one might have in mind. It is even more difficult to locate a usable listing of the sub-strata into which a given population might be divided. And, of course, proceeding without a list is not the solution: interviewing people on the street on any day of the week or hour of the day is very likely to give some segment of the local population a greater chance of being selected than others.

The basic principle of sampling can be restated as follows: There must be no logical connection between the method of sampling and the characteristic being sampled. Thus, using the corner house as a basis for sampling for income is a biased design, because living in a corner house is not independent of income. Of course, sampling on the basis of corner houses may be a perfectly random design if one is sampling for eye color, since there is no logical reason to believe that people living in corner houses are predominantly blue-eyed or brown-eyed. That is, there is no reason to suspect that blue-eyed people, for example, would be denied an equal chance of being selected if we used corner houses as the basis for our sampling. This design may be biased, however, if the variable under investigation is standing height, since there are indications of a correlation between standing height and vocational success and, therefore, income.

Whether or not a sample is a random sample cannot be determined by looking at it; a hand of thirteen cards of the

same suit, for example, can be dealt randomly. The criterion for randomness must be sought elsewhere—that is, in the process itself. A random sample can be defined as a sample which has been obtained by a random method. Such a sample would give results that approximate the true population value closer and closer as the sample size increases. The problem can be resolved at the operational level by superimposing a new characteristic—for example, the series of cardinal numbers—on the population, and by sampling in accordance with this new characteristic. The problem of sampling then becomes a matter of enumerating the population (where this is possible) and of selecting certain numbers at random, perhaps by means of a table of random numbers, such as that in Table 7-1. These numbers, by definition, constitute a random sample of the numbers assigned to the population, and, therefore, provide a correspondingly random sample of the population of individuals. The numbers selected in this way can be assumed to be independent of any characteristic; their use as the basis for sampling probably provides as valid a guarantee of randomness as it is possible to devise.

Another approach which is sometimes used is to take a *systematic* sample consisting of every *i*th member of the population. This can be done by deciding what fraction of the total population is to be included in the sample, and by taking every *i*th case in order to obtain a sample of the size required. Systematic sampling is generally an acceptable sampling procedure since, if one starts at random, it gives every individual in the population an equal chance of being included in the sample. It is, however, a faulty design when there is a cyclical pattern in the variable being investigated. For example, taking a traffic count from five to six o'clock in the afternoon of every day would obviously provide a biased estimate of the number of cars that go by a particular intersection during a given period. It is necessary to break the rhythm of the pattern in order to eliminate the bias it would promote. A systematic sample is also unacceptable when the variable is increasing rather rapidly. If the sample consists of 1 case out of every 100, it would make a difference if the sample included Cases 1, 101, 201, 301, and so on, or by contrast, Cases 99, 199, 299, 399, and so on.

One of the more questionable sampling practices is allow-

TABLE 7–1
Random Numbers (I)

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	56 50 26 71 07	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28
18 18 07 92 46	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75
62 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53
37 85 94 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98	35 85 29 48 39
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80
99 49 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79	94 77 24 21 90
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44	38 68 88 11 80
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76	15 54 55 95 52
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85	27 55 26 89 62
66 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66
14 90 84 45 11	75 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62	02 67 74 17 33
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74	94 44 67 16 94
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43
17 53 77 58 71	71 41 61 50 72	12 41 94 96 26	44 95 27 36 99	02 96 74 30 83
90 26 59 21 19	23 52 23 33 12	96 93 02 18 39	07 02 18 36 07	25 99 32 70 23
41 23 52 55 99	31 04 49 69 96	10 47 48 45 88	13 41 43 89 20	97 17 14 49 17
60 20 50 81 69	31 99 73 68 68	35 81 33 03 76	24 30 12 48 60	18 99 10 72 34
91 25 38 05 90	94 58 28 41 36	45 37 59 03 09	90 35 57 29 12	82 62 54 65 60
34 50 57 74 37	98 80 33 00 91	09 77 93 19 82	74 94 80 04 04	45 07 31 66 49
85 22 04 39 43	73 81 53 94 79	33 62 46 86 28	08 31 54 46 31	53 94 13 38 47
09 79 13 77 48	73 82 97 22 21	05 03 27 24 83	72 89 44 05 60	35 80 39 94 88
88 75 80 18 14	22 95 75 42 49	39 32 82 22 49	02 48 07 70 37	16 04 61 67 87
90 96 23 70 00	39 00 03 06 90	55 85 78 38 36	94 37 30 69 32	90 89 00 76 33

TABLE 7–1
Random Numbers (II)

53 74 23 99 67	61 32 28 69 84	94 62 67 86 24	98 33 41 19 95	47 53 53 38 09
63 38 06 86 54	99 00 65 26 94	02 82 90 23 07	79 62 67 80 60	75 91 12 81 19
35 30 58 21 46	06 72 17 10 94	25 21 31 75 96	49 28 24 00 49	55 65 79 78 07
63 43 36 82 69	65 51 18 37 88	61 38 44 12 45	32 92 85 88 65	54 34 81 85 35
98 25 37 55 26	01 91 82 81 46	74 71 12 94 97	24 02 71 37 07	03 92 18 66 75
02 63 21 17 69	71 50 80 89 56	38 15 70 11 48	43 40 45 86 98	00 83 26 91 03
64 55 22 21 82	48 22 28 06 00	61 54 13 43 91	82 78 12 23 29	06 66 24 12 27
85 07 26 13 89	01 10 07 82 04	59 63 69 36 03	69 11 15 83 80	13 29 54 19 28
58 54 16 24 15	51 54 44 82 00	62 61 65 04 69	38 18 65 18 97	85 72 13 49 21
34 85 27 84 87	61 48 64 56 26	90 18 48 13 26	37 70 15 42 57	65 65 80 39 07
03 92 18 27 46	57 99 16 96 56	30 33 72 85 22	84 64 38 56 98	99 01 30 98 64
62 95 30 27 59	37 75 41 66 48	86 97 80 61 45	23 53 04 01 63	45 76 08 64 27
08 45 93 15 22	60 21 75 46 91	98 77 27 85 42	28 88 61 08 84	69 62 03 42 73
07 08 55 18 40	45 44 75 13 90	24 94 96 61 02	57 55 66 83 15	73 42 37 11 61
01 85 89 95 66	51 10 19 34 88	15 84 97 19 75	12 76 39 43 78	64 63 91 08 25
72 84 71 14 35	19 11 58 49 26	50 11 17 17 76	86 31 57 20 18	95 60 78 46 75
88 78 28 16 84	13 52 53 94 53	75 45 69 30 96	73 89 65 70 31	99 17 43 48 76
45 17 75 65 57	28 40 19 72 12	25 12 74 75 67	60 40 60 81 19	24 62 01 61 16
96 76 28 12 54	22 01 11 94 25	71 96 16 16 88	68 64 36 74 45	19 59 50 88 92
43 31 67 72 30	24 02 94 08 63	38 32 36 66 02	69 36 38 25 39	48 03 45 15 22
50 44 66 44 21	66 06 58 05 62	68 15 54 35 02	42 35 48 96 32	14 52 41 52 48
22 66 22 15 86	26 63 75 41 99	58 42 36 72 24	58 37 52 18 51	03 37 18 39 11
96 24 40 14 51	23 22 30 88 57	95 67 47 29 83	94 69 40 06 07	18 16 36 78 86
31 73 91 61 19	60 20 72 93 48	98 57 07 23 69	65 95 39 69 58	56 80 30 19 44
78 60 73 99 84	43 89 94 36 45	56 69 47 07 41	90 22 91 07 12	78 35 34 08 72
84 37 90 61 56	70 10 23 98 05	85 11 34 76 60	76 48 45 34 60	01 64 18 39 96
36 67 10 08 23	98 93 35 08 86	99 29 76 29 81	33 34 91 58 93	63 14 52 32 52
07 28 59 07 48	89 64 58 89 75	83 85 62 27 89	30 14 78 56 27	86 63 59 80 02
10 15 83 87 60	79 24 31 66 56	21 48 24 06 93	91 98 94 05 49	01 47 59 38 00
55 19 68 97 65	03 73 52 16 56	00 53 55 90 27	33 42 29 38 87	22 13 88 83 34
53 81 29 13 39	35 01 20 71 34	62 33 74 82 14	53 73 19 09 03	56 54 29 56 93
51 86 32 68 92	33 98 74 66 99	40 14 71 94 58	45 94 19 38 81	14 44 99 81 07
35 91 70 29 13	80 03 54 07 27	96 94 78 32 66	50 95 52 74 33	13 80 55 62 54
37 71 67 95 13	20 02 44 95 94	64 85 04 05 72	01 32 90 76 14	53 89 74 60 41
93 66 13 83 27	92 79 64 64 72	28 54 96 53 84	48 14 52 98 94	56 07 93 89 30
02 96 08 45 65	13 05 00 41 84	93 07 54 72 59	21 45 57 09 77	19 48 56 27 44
49 83 43 48 35	82 88 33 69 96	72 36 04 19 76	47 45 15 18 60	82 11 08 95 97
84 60 71 62 46	40 80 81 30 37	34 39 23 05 38	25 15 35 71 30	88 12 57 21 77
18 17 30 88 71	44 91 14 88 47	89 23 30 63 15	56 34 20 47 89	99 82 93 24 98
79 69 10 61 78	71 32 76 95 62	87 00 22 58 40	92 54 01 75 25	43 11 71 99 31
75 93 36 57 83	56 20 14 82 11	74 21 97 90 65	96 42 68 63 86	74 54 13 26 94
38 30 92 29 03	06 28 81 39 38	62 25 06 84 63	61 29 08 93 67	04 32 92 08 09
51 29 50 10 34	31 57 75 95 80	51 97 02 74 77	76 15 48 49 44	18 55 63 77 09
21 31 38 86 24	37 79 81 53 74	73 24 16 10 33	52 83 90 94 76	70 47 14 54 36
29 01 23 87 88	58 02 39 37 67	42 10 14 20 92	16 55 23 42 45	54 96 09 11 06

From Ronald A. Fisher, and Frank Yates, *Statistical Tables for Biological, Agricultural, and Medical Research* (New York: Hafner, 1957).

ing the sample to select itself, as with a sample of letters to the editor or incomplete questionnaire results. Also faulty is the practice of allowing the interviewer—or some expert—to select the sample on the basis of judgment. While it is commonly used, especially in commercial polls, such an approach involves considerable risk, even if the investigator is trained and definite restrictions are imposed on his operations.

SAMPLING DESIGNS

Sampling designs range from the very elementary to elaborate designs, such as sequential and multi-stage sampling. No perfect or universally adequate sampling design has, as yet, been devised. The method to be used in a given investigation depends on the nature of the problem, the subjects to be located, the resources available, as well as on such factors as cost and administrative convenience. A pilot study can be valuable in saving time and expense, in uncovering potential sources of difficulty, and in providing the investigating staff with training both in statistical and in field work. Generally that method is best which gives the greatest degree of precision per unit of sampling cost, and the pilot study can help to obtain the values necessary for the derivation of the most effective design. It can be said that a sample is adequate when it is precise enough to allow the required confidence to be placed in the dependability of its results. More specifically, it is adequate when the standard error permits the bracketing of the population parameter within a band of precision sufficient to meet the requirements of the study.

Probability and Non-Probability Designs

Sampling designs can be classified into two broad categories: 1. *probability designs;* and 2. *non-probability designs.* In the former, randomness is the fundamental element of control. This was demonstrated in the binomial distribution based on the tossing of 10 coins a total of 1024 times. Empirical evidence has shown that such distributions duplicate themselves, time in and time out, with only slight variation. In such a situation, randomness causes the distribution to duplicate itself within random sampling errors as determined by formula. Such designs permit the specification of the precision that is ob-

tained, and the number of cases necessary to provide the required precision.

Non-probability designs, on the other hand, derive their control from the judgment of the investigator. For example, a pollster might be instructed to interview 100 persons passing a certain street corner, or to contact by phone so many storekeepers, so many housewives, or so many clerks. In non-probability sampling, the cases are selected on such bases as availability and interviewer judgment. Frequently, randomness is erroneously assumed to follow from the stratification of the population into relevant sub-populations. The advantage of non-probability designs lies largely in the area of convenience, which—along with the extra sample size sometimes possible for the same cost—is felt to compensate for the relative risk of possible bias. Commercial polls, for instance, frequently claim—with some degree of empirical justification—that the increase in the precision of probability sampling over non-probability sampling is too small to warrant the extra cost of a random sampling design, particularly since their present procedures are adequate for present demands. The fact that the actual selection is done by experienced field workers is obviously involved in their relative success. Frequently, on the other hand, such samples are over-weighted with the co-operative, the available, and so on. They depend too exclusively on uncontrolled factors and, especially, on the investigator's insight, and there is no statistical procedure permitting the determination of the margin of sampling errors.

Many people do not see the advantage of random sampling. It is their opinion that any person who has had vast experience in doing research can improve on chance in selecting a sample. This is, of course, erroneous. On the other hand, there are instances where the investigator does not want a representative sample—that is, he does not want a sample that represents any particular population. In some studies—such as exploratory surveys in which the object is to gain insight into the problem—the investigator may choose as his sample only informed persons who can provide him with the maximum degree of insight into his problem. In such cases, of course, what is required or expected is a wealth of ideas rather than simply a description of a given population.

At times, it may be possible to combine probability and non-probability sampling. This is, of course, a complicated procedure, which calls for a somewhat greater understanding of statistical procedures than the average student is likely to possess. It must also be remembered that complicated sampling designs are very frequently costly, and that greater precision may be obtained for a given cost by simply taking a larger sample of a standard variety.

Simple, unrestricted random sampling is the simplest (probability) sampling design, calling for nothing more than selecting the required number of cases at random from the specified population. This can be done by using a table of random numbers, a roulette, or any haphazard scheme—or even a systematic design. It is also the most fundamental, inasmuch as it underlies most of the more advanced designs.

Stratified Sampling Design

Stratified random sampling is a refinement of simple random sampling since, in addition to randomness, stratification introduces a secondary element of control as a means of increasing precision and representativeness. A stratified random sample is, in effect, a weighted combination of random subsamples joined to give an over-all sample value. For instance, if we were to study the weight of the adult residents of a given city, perhaps with a view to a possible air lift, all one would have to do would be to take a completely random sample of the people regardless of sex, obtain the average per capita weight, and multiply by the total number to be evacuated. However, since men tend to weigh more than women, an error might be introduced if, by chance, we were to pick more (or less) than the proportionate number of men. The likelihood of a sizable error from this source is relatively small, since the basic elements of randomness would keep the sex ratio of the sample relatively coincident with that of the population. In general, where a sufficiently large sample is taken, a simple random sampling can be depended on to provide a usable answer. On the other hand, somewhat greater precision might be obtained if we were to make sure that the number of men and of women in our sample was proportional to the men-women ratio in the population. This could be done easily if we knew the

town population consisted of 5500 women and 4500 men and we wished to take a sample of 100. We would simply take 55 women and 45 men. These could then be selected, completely at random, from the list of the residents—that is, 55 out of the 5500 women and 45 out of the 4500 men selected, perhaps by random numbers.

There may be times when it would be advantageous to take a disproportionate number of cases from the different strata. Sometimes the precision of a sample for a given cost can be increased by taking a smaller representation of the more homogeneous strata, and a larger sample of the more heterogeneous. Instead of taking a stratified sample in which the numbers in each of the strata are proportional to the number in the strata of the population, it is a rather common procedure, for example, to make the size of the sample per strata proportional to the product of the number and the standard deviation of the variable within each of the strata of the population. An even greater improvement over random sampling might be obtained if the sampling in each of the strata is made proportional to the product of the number and the standard deviation of the strata in the population, and inversely proportional to the square root of the cost per sampling unit in that particular strata. One can also take a larger sample from the more doubtful strata, and then weight the mean of each of the strata according to the proportionality in the population. This would be desirable, for instance, in connection with the electoral college for the election of the President, where greater precision for a given sample size and cost would be obtained by lightly sampling the "obvious" states and heavily sampling the "doubtful" ones.

The usual stratification factors are sex, age, socio-economic status, educational background, residence (urban or rural), and occupation. Other factors which might be involved in special issues include political-party affiliation, religion, and race. Stratified sampling is generally difficult to conduct inasmuch as we rarely have a usable listing of each of the strata. On the other hand, it is generally not necessary to stratify on multiple bases because the bases tend to be inter-correlated, and consequently, stratification on one or more of these factors will

generally result in relatively adequate stratification on a number of other related factors.

Stratified random sampling provides more precise results than simple random sampling only if stratification results in greater homogeneity within the strata, with respect to the trait under study, than would be found in the whole population taken as a unit. Then stratification is profitable, in the sense of giving more precise results, whenever the population can be broken down into sub-populations with characteristic differences with respect to the trait under investigation. In the problem regarding the air-lift, for example, one might profit from stratifying according to sex, since there is a characteristic sex difference with regard to weight; there would probably be no point in stratifying according to hair color, since this appears to be independent of weight. Furthermore, even if we did find average differences in weight of persons with different hair color, we would still have to determine the distribution of hair color in the population and, possibly, the variability involved so that we could get a weighted sum. This would, of course, increase the cost of our sample, and this would have to be balanced against the possibility of gaining more precise results by spending the same amount of money on getting a larger sample on the basis of simple random sampling.

Stratification is particularly appropriate in opinion polls where, on such issues as the appointment of Clare Boothe Luce as Ambassador to the Vatican, the expression of *strong approval, approval, undecided, disapproval, strong disapproval* might be related to such background factors as sex, political and religious affiliation, educational status, and so on. There would be no point in stratifying with respect to a variable which is presumably unrelated to the issue under study—for example, the month of birth. To be meaningful, the results of a study of this kind must be reported separately, according to strata, whenever characteristic differences exist among the subpopulations into which the population as a whole can be divided with respect to the issue under study.

It also must be pointed out that the stratification of the population according to such factors as sex is based on logical judgment or evidence of a characteristic difference. Once the

strata are set, however, sampling within each of the strata must be at random in line with the principles presented in the previous sections, and it is a serious error to assume that stratification, as such, removes the need for random selection within the strata. On the contrary, it is stratification that is not essential to good sampling, for the basic element of control in sampling is randomness; stratification simply provides a secondary control.

Purposive Sampling

Purposive sampling can be considered a form of stratified sampling in that the selection of the cases is governed by some criterion acting as a secondary control. At one end of the continuum, we have the type of probability sampling illustrated by the standardization of the Stanford-Binet, in which Terman and Merrill—on the premise that a correlation exists between socio-economic status and IQ, and that, therefore, any sample not representative of the population with respect to socio-economic status would also be suspect with respect to IQ—attempted to include a proportionate representation of each of the socio-economic strata of American society as revealed by the 1930 census. Thus, the characteristic of socio-economic status acted as a secondary control in the selection of the sample.

Another form of purposive sampling is *quota sampling,* which is also a form of stratified sampling except that, as the term is commonly used, it refers to a non-probability design in which the investigator, after having stratified his population, uses his judgment rather than randomness in selecting the cases within each of the strata. The results may be good or bad. In some instances, depending on the good sense as well as the good fortune of the investigator, the results may be as accurate as those obtained in probability sampling. Generally, however, such sampling is best used where the object is not to get precise statistics, but rather to collect typical opinions on a given issue. Quota sampling would be indicated in an exploratory study where the purpose is to develop insight so that later a more accurate study can be conducted with probability sampling. Quota sampling has advantages over probability sampling with respect to convenience. For instance, it permits the investigator to substitute one person for another in the case of a

refusal. This does not solve the problem of the bias connected with non-response, of course; it simply ignores it. According to the viewpoint expressed in this text, this represents over-confidence in the magic of having a given sample size, since, of course, ignoring bias is not the equivalent of solving it. Sometimes quota sampling takes the form of using a certain sample, and then trying to identify the population which it is supposed to represent. This is generally known as *populationing,* a procedure which, in view of its obvious limitations, is open to serious question.

Double Sampling

A rather frequent extension of the basic sampling design is multi-stage sampling, which is really a matter of sampling within samples. This might involve, for instance, sampling certain houses within certain blocks of a given city, or certain classrooms within certain schools of the state or the system. Another example might involve the interviewing of non-respondents to a questionnaire to determine the nature of the reactions of that particular segment of the over-all sample, and the weighting of their responses in order to give them fair representation in the final results of the total sample. It must, of course, be noted that double sampling complicates the statistical analysis of the data and correspondingly increases the cost.

In the usual double sampling design, the investigator selects his sample on the basis of a characteristic which is readily available and highly correlated with the primary characteristic for which the collection of data is expensive and/or difficult. Since the two characteristics are correlated, an adequate sample with respect to the second characteristic should automatically also be an adequate sample with respect to the first.

An interesting variation of double sampling occurs when the values of the primary characteristic are obtained by means of an equation relating it to a secondary characteristic for which an adequate sample can be obtained. For example, one might want to determine the amount of money teachers on regular contract contribute to local grocery stores and restaurants. A random sample of teachers, perhaps stratified according to marital status and other relevant factors, could be obtained. It would then be a matter of asking each how much he

spends for food, getting an average figure for each of the strata, and getting the grand sum. However, some teachers do not keep such records; asking them to keep records for the next month would probably make them money-and-food-conscious, and promote a completely distorted view. Thus, an investigator cannot proceed with such non-record-keepers, nor without them. A scheme which is also not completely free from flaws, but which might be relatively accurate, is to proceed indirectly through devising an equation for teachers who keep accounts, relating food expenditures to such factors as salary, family size, and so on. For example, a crude equation might be

$$F = .06S\sqrt{2n} + \$35$$

According to this equation, a teacher earning about $500 a month with a family of two children (plus two adults) would spend:

$$\begin{aligned} F &= .06(\$500)\sqrt{2(4)} + \$35 \\ &= \$119 \end{aligned}$$

Assuming such an equation to be fairly adequate, and assuming further that teachers who do not keep accounts have eating habits not radically different from those who do, we can get the average food-cost for teachers by substituting in the equation the average salary and the average family size, and multiplying by the number of teachers in the system. If necessary, separate equations could be devised for each strata and a weighted sum obtained.

Cluster Sampling

A fundamental problem connected with sampling concerns the choice of the sampling unit. Although generally the sample is selected in units of one, this need not be so, especially in education, where it is frequently as easy to contact a whole classroom as it is to contact a single individual. This sampling design in which the unit of sampling consists of multiple cases—for example, a family, a classroom, a school, or even a city or a school system—is known as *cluster sampling*. Thus, in the standardization of the Stanford-Binet, Terman and Merrill selected a given community and tested every single child in that community who was within one month of his birthday.

Cluster sampling is particularly attractive from the standpoint of permitting the easy accumulation of large samples. This is, however, somewhat misleading in that, to the extent that the members of individual clusters are more homogeneous than an equal number of cases selected completely at random (that is, to the extent that a positive intra-class correlation exists among the members), an overlapping effect takes place so that the effective number of cases from the standpoint of increasing the precision of the sample is somewhat less than the actual number of cases included.[5]

Nevertheless, even if a substantial intra-class correlation exists, a cluster sampling design generally is advantageous in that the loss of precision per individual case is more than compensated for by the possibility of taking larger samples for the same cost. It is agreed, however, that a sample obtained by taking a relatively large number of small clusters is preferable to a sample of equal size obtained on the basis of a small number of large clusters.

Cluster sampling is independent of the other kinds and classifications of sampling designs, and one might sample in clusters according to a simple random sampling design, a stratified random sampling design, or any other sampling design. For example, in a study of high-school seniors the sampling unit might be the English class; each English class in the state can be numbered; stratification can be made according to the size of the school; then, by means of random numbers certain English classes can be selected, and tested as a unit.

Sequential Sampling

An interesting sampling design of rather recent origin is *sequential sampling* in which sampling is continued until a significant result on which to base a decision is obtained. For instance, a manufacturer having devised a new light bulb would want to test this bulb for life expectancy before placing

[5] The computation of the standard error in cluster sampling calls for a special formula which is somewhat more involved than that for the single sampling unit, especially when inequality in the size of the cluster is found. See Russell L. Ackoff, *The Design of Social Research* (Chicago: University of Chicago Press, 1953), p. 114; Leon Festinger and Daniel Katz, *Research Methods in the Behavioral Sciences* (New York: Dryden, 1953), p. 203ff.; or Eli S. Marks, "Sampling in the Revision of the Stanford-Binet Scale," *Psychological Bulletin*, 44 (September, 1947): 413–34.

it on the market. Since testing the bulb would imply its destruction, however, he would want to conduct the test as economically as possible. This he might do by testing, perhaps, fifty bulbs. If these proved to be significantly superior or significantly inferior to the conventional bulbs, he would then have his answer. If, however, the test proved to be inconclusive, he would then have to add another fifty bulbs for an overall test of one hundred bulbs. This might provide a conclusive answer; if not, the test would be continued by the addition of one batch of fifty bulbs after another until the issue is settled one way or the other—and at a minimum expense.

Sequential sampling introduces an interesting approach to research. Thus, instead of carrying out a study of five hundred cases, it might be advisable to carry, say, a five-stage sequential research program of one hundred cases each. If the first step provides a decisive answer, the study can be dropped immediately. If not, it can be continued until the answer is obtained, or until the five hundred cases are exhausted. In such an approach, if a basic flaw were to be noted in the design of the study, the first stage could be considered a pilot study to the others, which would then be conducted on the basis of an improved design.

Synthesis

In summary, it might be repeated that there is no best sampling design; validity of sample data, like validity of all data, is a specific concept to be evaluated from the standpoint of the specific case. It is, therefore, difficult to generalize. Nevertheless, it generally is true that the aspect of sampling to which investigators of educational problems might most profitably devote their attention is minimizing possible bias, rather than devising complicated designs.

SUMMARY

1. Research is invariably conducted on the basis of a sample on the basis of which inferences concerning the population can be derived through statistical procedures. Sampling is both necessary and advantageous in the usual case. It is especially fundamental in survey research.

2. If a sample is to serve as the basis for inferences concerning the population, it is essential that it be representative—that is, that

it be a replica—of the population in question. Since this principle is impossible to implement, statisticians have substituted the concept of randomness with the understanding that a random sample will provide statistics within random sampling errors of the corresponding population parameters. The magnitude of these errors can be estimated at any probability level, and the population parameters can therefore be estimated on the basis of probability to lie within specified intervals.

3. A sample that is not representative can suffer from errors of a random and/or systematic nature and further from errors of sampling and/or measurement. Random errors of both sampling and measurement can be reduced to fractional values—even to the point of complete elimination—by increasing the sample size. Not only can their magnitude be estimated, but the size of the sample necessary to provide a desired degree of precision at a given probability level can be computed in advance if the sampling distribution of the statistic is known. Random errors can also be reduced through an improved sampling design. Constant errors, on the other hand, are simply stabilized (rather than eliminated) by taking even a substantial sample. Constant errors of measurement are not removed even by taking a complete census.

4. The first problem in sampling is to define the population so that there is no doubt about who is to be included and to whom the results of the study are to apply.

5. A basic principle of sampling is that every member of the population must have an equal chance of being included in the sample. This immediately raises the complication that it is almost impossible to obtain an adequate listing of any population from which the sample might be selected. A somewhat more readily applicable principle of sampling is that there must be no logical connection between the method of sampling and the characteristic being sampled. Where the population can be enumerated, this principle is generally best implemented through the use of a table of random numbers.

6. Sampling can be based on a probability or a non-probability design. The latter derives its control from the judgment of the investigator, not only is it subject to serious error, but it does not provide the basis for calculating the magnitude of such error. Probability designs, on the contrary, derive their control from the concept of randomness and thus, can provide an appraisal of random errors.

7. The basic sampling design is simple random sampling. Stratified sampling introduces a secondary control and provides greater precision in sampling whenever stratification results in greater homogeneity in the substrata with respect to the variable in question.

8. Cluster sampling is of interest to educational researchers who can frequently select their samples in units of a classroom as easily as in units of a single child.

9. A number of other sampling designs are possible, some of which are relatively complicated from the standpoint of both sampling and statistical treatment. In general, educational researchers might more profitably orient their efforts to minimizing possible biases in sampling and measurement than to experimenting with complex sampling designs.

SELECTED REFERENCES

ARRINGTON, RUTH E. "Time Sampling in Studies of Social Behavior; A Review of Techniques and Results with Research Suggestions," *Psychological Bulletin*, 40 (February 1943) : 81–124.

CANTRIL, HADLEY. *Gauging Public Opinion*. Princeton: Princeton University Press, 1947.

COCHRAN, WILLIAM G. "Modern Methods in the Sampling of Human Populations," *American Journal of Public Health*, 41 (June 1951) : 647–53.

———. *Sampling Techniques*. New York: Wiley, 1953.

———. "Design and Analysis of Samplings," in Snedecor, George W., *Statistical Methods*. Ames: Iowa State College, 1956.

COCHRAN, WILLIAM G., *et al.* "Principles of Sampling," *Journal of the American Statistical Association*, 49 (March 1954) : 13–35.

DEMING, WILLIAM E. "On Errors in Surveys," *American Sociological Review*, 9 (August 1944) : 359–69.

———. *Some Theory of Sampling*. New York: Wiley, 1950.

FISKE, DONALD W. and JONES, LYLE V. "Sequential Analysis in Psychological Research," *Psychological Bulletin*, 51 (May 1954) : 260–75.

FOTHERINGHAM, WALLACE C. "Drawing Highly Representative Samples from Large Populations," *Educational Research Bulletin*, 34 (April 1955) : 95–7.

GOODMAN, ROE and KISH, LESLIE. "Controlled Selection—A Technique in Probability Sampling," *Journal of the American Statistical Association*, 45 (September 1950) : 350–72.

HANSEN, MORRIS H. "Modern Methods in the Sampling of Human Populations," *American Journal of Public Health*, 41 (June 1951) : 662–8.

HANSEN, MORRIS H. and HURWITZ, WILLIAM N. "The Problem of Non-Response of Sample Surveys," *Journal of the American Statistical Association*, 41 (December 1946) : 517–29.

HANSEN, MORRIS H., *et al. Sample Survey Methods and Theory. I: Methods and Applications, II: Theory*. New York: Wiley, 1953.

HARTLEY, H. O. and YATES, FRANK. "Review of Recent Statistical Developments in Sampling and Sampling Surveys; Methods of Estimating Errors with Discussion," *Journal of the Royal Statistical Society*, 109 (1946) : 12–43.

KENDALL, MAURICE G. and SMITH, B. BABINGTON. "Randomness and Random Sampling Numbers," *Journal of the Royal Statistical Society*, 101 (1938) : 147–56.

———. "Second Paper on Random Sampling Numbers," *Journal of the Royal Statistical Society, Suppl.* 6 (1939): 51–61.
———. *Table of Random Sampling Numbers,* Tract for Computers. No. 24. London: Cambridge University Press, 1951.
LINDQUIST, EVERETT F. "Sampling in Educational Research," *Journal of Educational Psychology,* 31 (November 1940): 561–74.
MCCARTHY, PHILIP J. *Sampling: Elementary Principles.* Bulletin No. 15. Ithaca: Cornell University, 1951.
MCNEMAR, QUINN. "Sampling in Psychological Research," *Psychological Bulletin,* 37 (June 1940) 331–65.
MARKS, ELI S. "Selective Sampling in Psychological Research," *Psychological Bulletin,* 44 (May 1947): 267–75.
———. "Sampling in the Revision of the Stanford-Binet Scale," *Psychological Bulletin,* 44 (September 1947): 413–34.
———. "Some Sampling Problems in Educational Research," *Journal of Educational Psychology,* 42 (February 1951): 85–96.
MEIER, NORMAN C. and BURKE, CLETUS J. "Laboratory Tests of Sampling Techniques," *Public Opinion Quarterly,* 11 (Winter 1947): 586–95; 12 (Spring 1948): 316–24.
MOONAN, WILLIAM J. "Some Empirical Aspects of the Sequential Analysis Technique as Applied to an Achievement Examination," *Journal of Experimental Education,* 18 (March 1950): 195–207
MYERS, ROBERT J. "Accuracy of Age Reporting in the 1950 United States Census," *Journal of the American Statistical Association,* 49 (December 1954): 826–31.
RAO, C. RADHAKRISHNA. "Sequential Tests of the Null Hypothesis," *Sankhya,* 10 (November 1950): 3161–9.
SLONIN, MORRIS J. "Sampling in a Nutshell," *Journal of the American Statistical Association,* 52 (June 1957): 143–61.
STEPHAN, FREDERICK F. "History of the Uses of Modern Sampling Procedures," *Journal of the American Statistical Association,* 43 (November 1948): 12–39.
STEPHAN, FREDERICK F. and MCCARTHY, PHILIP J. *Sampling Opinions: An Analysis of Survey Procedures.* New York: Wiley, 1958.
SUKHATME, PANDURANG V. *Sampling: Theory of Survey.* Ames: Iowa State College Press, 1954.
TUKEY, JOHN W. "Some Sampling Simplified," *Journal of the American Statistical Association,* 45 (December 1950): 501–19.
UNITED STATES DEPARTMENT OF COMMERCE, BUREAU OF THE CENSUS. *Illustrative Projections of the Population of the United States, by Age and Sex: 1955–75.* Current Population Reports, Series P-25, No. 78. Washington: Bureau of the Census, 1953. *Revised Projections of the Population of the United States, by Age and Sex, 1960–75.* Series P-25, No. 123, 1955.
WALKER, HELEN M. "Educational Research and Poll-Taking," *School Management,* 18 (June 1949): 9. (See also *High Points,* 31 (January 1949): 5–7.)

PART III: RESEARCH METHODS

8 Introduction to Research Methods

Classification is inevitably an arbitrary process, resulting in a product of varying degrees of appropriateness and usefulness depending on the nature of the phenomena to be classified and the purpose to be served. The categorizing of educational research methods into logical and functional classes is doubly precarious because of the composite and overlapping nature of many of its procedures. Yet, despite this lack of clear-cut distinctions among the methods, it is desirable to attempt their classification for the insights into the overall organization and nature of educational research which such attempts provide.

That there is no natural system of classification of educational research methods which would cause each of the methods to fall neatly into place becomes evident when one considers the differences in the classification systems presented by the different authors of textbooks and articles in the field. As Barr points out, educational research methods can be categorized on the basis of end result (or goal), data-gathering technique, method of data-processing, degree of control exercised, approach, source of the data, and a number of other considerations. Educational research can also be classified as laboratory or field research, action or pure research, and, of course, according to such other dimensions as curriculum research, psychometric research, or sociometric research.[1]

[1] Arvil S. Barr, "Research Methods," in Chester W. Harris (ed.), *Encyclopedia of Educational Research* (New York: Macmillan, 1960), pp. 1160–6.

In practice, most authors agree on three basic categories:

1. *Historical*, which is concerned with the past and which attempts to trace the past as a means of seeing the present in perspective.
2. *Survey*, which is concerned with the present and attempts to determine the status of the phenomenon under investigation.
3. *Experimental*, which is oriented toward the discovery of basic relationships among phenomena as a means of predicting and, eventually, controlling their occurrence.

This classification is based partially on time sequence though, to be sure, even more significant differences also exist with respect to the purposes which the methods are to serve, the nature of the problems for which they are appropriate, and the procedures employed in the conduct of each.

This basic classification is used by Best[2] in his text. Hillway[3] adds a fourth category—the case study. Good and Scates[4] also add a fourth category to cover the area of complex causal relationships. More specifically, they add research of a causal-comparative, correlational, case study, and genetic nature. Travers,[5] on the other hand, follows a somewhat different organization; he omits historical research, on the grounds that it is relatively impossible to derive historical data suitable for the testing of hypotheses. Cornell and Monroe[6] also present a more complex system of classification: not only do they list five basic classes—descriptive, metric, clinical, correlational, and experimental—but they also mention as a possible sixth method, "theory construction or model building and the verification of theoretical systems."

The discussion of research methods in the present text will be organized according to the three basic categories outlined above. More specifically, the various educational research methods will be considered under the following headings:

[2] John W. Best, *Research in Education* (Englewood Cliffs: Prentice Hall, 1959).

[3] Tyrus Hillway, *Introduction to Research* (Boston: Houghton-Mifflin, 1956).

[4] Carter V. Good and Douglas E. Scates, *Methods of Research* (New York: Appleton-Century-Crofts, 1954).

[5] Robert W. M. Travers, *An Introduction to Educational Research* (New York: Macmillan, 1958).

[6] Frances G. Cornell and Walter S. Monroe, "Productive Methods in Research," *Phi Delta Kappan*, 35 (October 1953): 29–34.

Historical
1. historical
2. legal
3. documentary

Survey
1. descriptive
   a) survey testing
   b) questionnaire
   c) interview
2. analytical
   a) documentary-frequency
   b) observational
   c) rating
   d) critical incident
   e) factor analysis
3. school surveys
4. social surveys
5. genetic

Experimental
1. simple experimental designs
2. multivariate analysis
3. case study
4. predictive (correlational)

The distinction between the various categories is, of course, imprecise, and the reader might be tempted to question the specific allocation of certain kinds of research to the particular category to which they have been assigned. From the standpoint of purpose—that is, determining the status of a given phenomenon—legal and documentary research are, for example, perhaps more closely related to survey than to historical research. On the other hand, the particular problems encountered, and the specific techniques to be applied, in such research probably more closely resemble those of historical research, and, for the sake of organization, reader comprehension, and the avoidance of unnecessary repetition, they are discussed in that setting.

The particular allocation of the various methods to a category is essentially a matter of judgment, and the classification of the different methods here is primarily a scheme for unified presentation rather than a rigid, mutually-exclusive organiza-

tion which is inherent in the different methods. In fact, though there are basic similarities in the methods grouped in each category, at times, there are also considerable differences. Actually, no two problems can be solved in identically the same way; what constitutes the proper method for dealing with a specific problem can be decided only on the basis of its pecularities. Furthermore, what is relatively the same analysis of essentially the same data might fall in one category or another depending on one's purpose, and, of course, a given method is frequently used in a subsidiary way in conducting research based on another classification—for example, interviewing as a means of dealing with non-response in a questionnaire study.

No one system of classification can fit a field as complex as education. On the contrary, if they are to be effective in dealing with problems of the complexity of those in education, educational research methods must be varied, complex, and, inevitably, overlapping. This is especially true inasmuch as, at the present stage of its development as a science, education needs exploratory studies that have general significance in broad general areas. Later, as the field becomes more clearly defined, it will become progressively more possible and more necessary to emphasize controlled experimentation.

In a sense, it is relatively futile even to concentrate on the identification of research methods according to a rigid categorization. Our efforts might be more profitably directed toward seeing that the method used is in harmony with scientific principles, and that it is adequate for the job. Conversely, any method, or any combination of methods, that leads to dependable generalization is automatically a good method. There is, however, a need to define and to evaluate the method used, and, as Hillway[7] points out, if one cannot describe his approach, chances are that his understanding of what he is doing is too vague and that his approach will prove ineffective. There is also the need for a thorough understanding of all research methods—with particular reference to their strengths, limitations, applicability, and appropriateness—for an inappropriate method can only lead to unsatisfactory results and disillusionment.

It is worthwhile to repeat that, while the methods listed

[7] Tyrus Hillway, *op. cit.*, p. 126.

entail obvious differences in purpose and approach, the significant aspect of the situation is their similarity as techniques of science, for despite their superficial differences, they qualify as research methods only as they adhere to the basic principles of science and scholarliness. This is demonstrated—in all research methods—in the precision with which the problem is formulated, the population defined, and the sample selected; the care with which the data are collected, validated, and interpreted; and the scholarship with which inferences are drawn and the report is written. It is only within the framework of this basic similarity that their differences exist.

Man is the only creature who is aware of and interested in his past.

JAMES W. THOMPSON

9 Historical Research

Historical research is one of the most difficult types of investigation to conduct adequately. Although everyone is a historian in that he remembers what occurred in the past, such "history" does not meet the criteria of historical research which, if it is to be a science, must meet the same standards of excellence as other forms of research.

NATURE OF HISTORICAL RESEARCH

The term *history* is variously used and in order to place historical research in its proper perspective, a brief overview or general orientation to the nature and development of history will be presented. As used by the early Greeks, history meant an inquiry to establish what had actually happened, and, to some degree, history is still that branch of learning that studies and records past events. As it applies to research, history is first of all, an inquiry, an attempt to discover what has happened. To historians of the later nineteenth century, this was the only function of the historian. They also believed that, by subscription to a scientific historical approach—reliance on dependable sources, the authentication of sources, and the validation of evidence through an elaborate system of internal criticism, together with as complete an objectivity as humanly possible—the past could actually be discovered.

While some of these considerations are still valid today, few modern historians would be bold enough to claim that complete objectivity is possible—or perhaps even desirable. Nor do they ever hope to discover the past as it actually happened. They no longer conceive of their function as that of a simple recorder of past events concerned with the establishment of facts. More and more the emphasis has been toward the interpretation of the data, toward giving meaning to the events described rather than simply producing an encyclopedic catalogue of events. Here the historian is on less sure grounds, and he must be sure that his conclusions are based on as verifiable data as he can gather; it is here that the historian stakes his claim to scholarliness.

The historian is inevitably influenced by some philosophy operating explicitly or implicitly in his interpretations. Historical philosophies generally fall into three major categories. The first sees history as an expression of a plan or purpose set by divine or natural (scientific) law which simply leads man on to his destiny. The second views civilization as a biological organism with the determinants of its developments, its achievements, and its life span inborn. Finally, there is the humanistic view, which gives man an important role in determining his fate and that of the world of which he is a part. The historian generally espouses a theoretical position and attempts to interpret his data with respect to the broad theoretical positions listed above, or to some of the more specific theories, such as the scientific (technological), economic, geographic, great-man, or even the eclectic theory, in order to give the facts of history meaning.

Historical research can be classified according to 1. *approach*—for example, the pragmatic approach used by Karl Marx to arrange all the facts of history to support his concept of socialism; 2. *subject*—for example, the biography of a given person, the monography of a town, state, nation, or a civilization, or, at a slightly higher level, the history of ideas, institutions, or trends; and 3. *technique*—that is, based either on documents or on relics.

Purposes of Historical Research

The purposes for which historical research is undertaken are probably as varied as the many individuals who engage in

the activity. They can, however, be summarized under two major heads:

1. The foremost purpose of doing historical research is to gain a clearer perspective of the present. Present problems—for instance, the current opposition to federal aid to education or racial integration and segregation—are understandable only on the basis of their past. Most things have a history, and it is generally profitable to acquaint ourselves with this history if we are really to appreciate their nature. Historical research can provide us not only with hypotheses for the solution of current problems, but also with a greater appreciation of the culture and of the role which education is to play in the progress of society.

An understanding of the historical background of education should enable the educator to recognize fads and frills, which are frequently advocated as the "just discovered" cure for educational ills, when in reality, they are simply rejuvenated versions of ideas tried years ago and found to be wanting. This does not mean that these ideas are not to be reconsidered, since changes in the interim may have put them in a new light, but it should still be noted that they are not new. Stiff grading, for example, is not something that was spawned by Sputnik I, nor is the four-quarter school year an invention of the 1960's—though new developments may have placed these ideas in different perspective. An understanding of its historical background should save education from making the same mistakes. Thus historical research can act as a control in policy-making.

2. A common motive underlying historical research is the simple scholarly desire of the scientist to arrive at an accurate account of the past. This may involve nothing more than a scholarly interest in truth—that is, the desire to know what happened in the past, and how, and "why the men of the times allowed it to happen."[1] There is even room for the scientist to be interested in giving an accurate account of the past without particular concern for its meaning for the present. On the other hand, the historian generally would not be satisfied with the mere discovery of truth, but would conceive his primary

[1] Henry L. Smith and Johnnie R. Smith, *An Introduction to Research in Education* (Bloomington's Educational Publications, 1959) p. 127.

responsibility as a scientist to be the interpretation of the data in order to link the past to the present and to the future.

The Steps of Historical Research

Although slight adaptations from standard scientific methods need to be made because of its nature, historical research must meet the same criteria and generally follow the same procedures as the other forms of research.

1. The identification and delineation of the problem is frequently a difficult proposition, since it involves not only the location of a problem which has historical and current significance, but also the availability of adequate data. Many otherwise acceptable historical topics may have to be discarded when data simply are not available. Thus it might be a nice problem to determine conclusively the authorship of the Shakesperian plays, but probably little could be located that would add any light to the present uncertainty.
2. The collection of data may involve anything from digging up ancient ruins to chancing on old documents, such as the Dead Sea Scrolls. Although materials occasionally may be found in old manuscripts located by chance, most educational data probably have to be located in the routine fashion of going through minutes of meetings, diaries, and so on.
3. The establishment of the validity of the data generally involves the dual process of first establishing the authenticity of the source, and then the validity of its contents.
4. The interpretation of the data must be made from the standpoint of whatever hypothesis or theory the data will most adequately support. Isolated facts have no meaning, and a mere listing of historical occurrences is not research. It is necessary that data be considered in relation to one another and synthesized into a generalization or conclusion which places their overall significance in focus.

HISTORICAL EVIDENCE

The Nature of Historical Evidence

The difficulty of deriving truth from historical evidence, and the methodical care used by historians in dealing with this fully recognized problem, must be realized if historical science is to be properly appreciated. The major problem is, of

course, that the data on which such research is based are invariably relatively inadequate. Usually at the time a historical study is conducted the sources are no longer available for complete investigation, and consequently much of the data has to be inferred—with all of the undependability this may entail. Lack of perspective, as well as lack of impartiality and disinterestedness, can, of course, make it equally difficult to deal with more current events. Thus, in appraising the tons of data gathered during World War II, it would be difficult to maintain complete objectivity, especially with regard to events in which people are emotionally involved or that concern persons who are still living. It is quite likely that embarrassing or incriminating, as well as confidential, information will be suppressed, and that our failures and successes will undergo some degree of distortion.

The date of the occurrence of a given event often is difficult to determine, partly because of the confusion arising from the change to our present calendar. The calendar was revised in the sixteenth century by Pope Gregory XIII, but the revision was not accepted in British countries until the mid-eighteenth century, at which time the current calendar was some eleven days behind the new Gregorian calendar. As a result, we recognize Washington's birthdate as February 22, while according to the family Bible his birthdate was February 11. There was no year 0 in the new calendar; it merely skipped from 1 B.C. to 1 A.D. Furthermore, there are reasons to believe that Jesus Christ was born in 7 B.C., 4 A.D., or even 6 A.D., rather than in the year 1.

It is frequently difficult to determine the date when a certain university was established, for it may have operated on a semi-organized basis, or at a lower level of education, for a number of years before it became fully chartered as an institution of higher learning. Or it may not have had any students or even its own physical plant for a number of years even after the charter was granted. Brickman[2] gives instances of a degree-granting college arbitrarily choosing as its official opening date, the date on which it first began as an elementary school.

The term *first* is also troublesome. For instance, the first

[2] William W. Brickman, *Guide to Research in Educational History* (New York: New York University Bookstore, 1949).

psychological laboratory in America is variously credited to J. McKeen Cattell, and to William James, depending on whether keeping experimental animals in the basement constitutes a laboratory. Similarly, which school qualifies as the first normal school depends on whether we are thinking of the first school to perform the functions of a normal school, or whether we insist that it be chartered by the state under that title.

Sources of Historical Evidence

Historical sources may be classified in two major categories —*documents* and *relics* (or remains) —according to whether or not the source was designed specifically for transmitting information, or whether it is simply an artifact. Documents are usually written, whereas relics, since they are generally archaeological or geological remains like tools and utensils, are usually unwritten—but this is not the basic point of distinction. A letter written by Lincoln, for example, would be a document from the standpoint of the information it contains but would be a relic from the standpoint of spelling errors or other aspects not part of what Lincoln intended to transmit.

Among the various documentary sources we may list 1. official records—minutes of meetings, committee reports and legal documents; 2. institutional records—attendance rolls, university bulletins, and so on; 3. memoirs, biographies, diaries, personal letters, books on the philosophy of a given scholar, and so on. There are, of course, a number of limitations inherent in each of these sources. In the wording of laws that are finally passed, for instance, there is a suggestion of unanimous agreement though violent discussions may have preceded their acceptance, many modifications and amendments may have been suggested, until, eventually, a compromise was reached—which may be to no one's liking. The policies listed in university bulletins frequently are nullified through numerous exceptions. In the memoirs of faculty members, occurrences of many years past frequently take on a new light as the author sees his career in perspective. Manuscripts are frequently subjected to so many editorial changes that they no longer resemble their original form. Newspaper articles of educational events are particularly subject to distortion, either

through careless reporting or through emphasis on the sensational, with a corresponding complete disregard of the educationally significant aspects of the situation.

Primary and Secondary Sources

The historian's first step in evaluating the adequacy of his evidence is to distinguish between evidence from *primary* sources—that is, data provided by actual witnesses to the incident in question—and evidence from *secondary* sources in which a middleman has come between the original witness and the present consumer. Secondary sources are subject to an inherent danger of inaccuracy; whenever evidence is transmitted from one person to another it tends to become distorted. Occasionally secondary sources have been so carelessly compiled that they are in the category of unverified hearsay or rumor. For this reason, reliable historians rely as much as possible on primary sources, using secondary sources only as hypotheses to bridge the gaps between the various pieces of primary evidence.

It is not always possible, of course, to obtain primary evidence, and at times the historian may have to rely on secondary sources. He must be fully aware of the limitations of such data, however, and, in the event that numerous gaps in the primary sources cause his over-reliance on secondary sources, he should probably refrain from attempting the study at all. This is a common problem in education where, surprisingly enough, only fragmentary reports concerning the processes of education are available. It seems that people in the past considered education so fundamental and so commonplace that they did not bother recording anything about its nature or its organization. Consequently, it is relatively difficult to locate suitable evidence to permit the conduct of a good historical study in education. Such personal documents as diaries and personal letters also leave too many gaps for the average historian to get the required continuity, without undue resorting to secondary sources and his own imagination.

On the other hand, though it is true that frequently what is called history is so far removed from the original source and so carelessly compiled that it is unacceptable, it is also true that secondary sources are sometimes particularly accurate. If the historian has an adequate insight into the situation so

that he can balance one secondary source against another, he may come much closer to the truth than he would if he relied on a single original source. For instance, we frequently listen to news commentators for an orientation to a message from the President, for commentators, because of their backgrounds, frequently are able to synthesize the significant factors in the situation and present a much clearer picture than can be obtained from a first-hand report. For the same reason, it may be better to read a translation of a given passage than to read the original in a foreign language. It should also be understood that secondary sources often become more accurate with the passage of time, as historians gain impartiality as well as perspective and, of course, as more data become available. The historian, therefore, does not ignore or reject secondary sources; he investigates any lead he can uncover, but he does not believe anything until he has investigated its validity. In fact, while the historian uses both primary and secondary sources as the bases for hypotheses, he subjects both to rigorous tests.

Criticism of Historical Data

Historians are fully aware of the limitations of the data with which they have to deal and have developed systematic means of evaluating such evidence. Generally, the criticism of historical data involves the dual processes of establishing the authenticity of the source and of establishing the validity of its contents. These are known as *external* and *internal* criticism, respectively, though the terms *lower* and *higher* criticism are also used.

Establishing the Authorship of Historical Data

In evaluating evidence, the historian must first establish the authenticity of his sources. Stated negatively, he must save himself from being the victim of a fraud such as those which have at times been perpetrated not only on the public but even on scientists. A classic example of a fraud, uncovered in recent times, is the Cardiff Giant, which was presumably the fossilized skeleton of a pre-historic human monster found near Syracuse in 1869. Cleverly made out of gypsum at the request of a farmer and buried in a spot where it would be found by well-diggers, it was acclaimed, even by some reputable scientists, as

genuine. The fraud was finally exposed by a newspaper reporter who was able to establish that a shipment of gypsum had been made to a certain barn prior to the discovery of the skeleton, and who secured a confession from one of the participants.

Although few students in education are likely to encounter major problems in establishing the authenticity of documents and other evidence, it is important for them to realize that very elaborate techniques have been developed to forestall the perpetration of frauds and forgeries. These range from the application of logic to the use of the most sensitive devices of modern science. For example, the philologist may be able to detect frauds on the basis of the change in the meaning of words over the years. Forgeries have been exposed through their use of words that were not "invented" or that were no longer in use at the time the document is alleged to have been written. The improper use of historical time, such as Lincoln's alleged reference to the state of Kansas at a time when the state did not exist, is a slip that can be detected through the science of anachronism. Frauds in documents can sometimes be detected by dating the document by chemical analysis of the paper, the watermark, the ink, or investigation of the spelling and the language in use at the alleged time of publication. The fluorine and the Carbon-14 tests can be used to establish the age of fossils and other remains. Recently, ultra-violet rays and fluorescent photography have been developed as a means of detecting erasures and alterations. Many other means of detecting frauds could, of course, be mentioned. It is, however, likely that the best tool in the detection of frauds is the investigator's common sense combined with a healthy sprinkling of skepticism.

The purpose of external criticism is, however, not so much negative—that is, the detection of fraud—as it is the establishment of historical truth. Of course, external criticism, though capable of causing the rejection of a fraud, cannot "prove" the genuineness of a particular source except on the basis of plausibility and probability.

Among the more common problems encountered in establishing the authenticity of documents are those of plagiarism and alterations. For example, after a speech is given and before it is made part of the record, it may be reviewed for gram-

matical errors and, in some cases, for alteration of content. Such modifications can be found even in the Congressional Record (which incidentally also contains records of "speeches" that were never given on the floor). The relatively common practice of having reports ghost-written also causes difficulties. It is felt, for instance, that Washington's inaugural address was written by Jefferson. In fact, a person in high office may incorporate in a report material compiled by different subordinates —perhaps with different styles of writing—which might lead later historians to suspect a forgery or an alteration of the original document.

Establishing the Validity of Historical Data

Even more crucial from the standpoint of the basic purpose of science—the derivation of truth—is the establishment of the validity of the content of a document or source, regardless of its authorship or genuineness. This is frequently no simple achievement.

For example, changes often have been made in older documents as a result of errors, omissions, additions, and transpositions in transcribing. Before the advent of printing, each document had to be copied by hand, and, though much of this was done with particular care by monks and other scholars, occasional errors crept in. To make matters worse, these documents frequently were copied from one copy to next so that the document now available may be many times removed from the original—which may have been lost or simply destroyed. For example, we do not have an original copy of Chaucer's tales. Furthermore, as the documents were copied over the centuries, interpretative notes placed in the margin by one scholar may have been included as part of the regular text when it was copied by the next scholar, so that the final copy, though genuine as a whole, may have any number of parts that were not in the original.

In addition to copying errors, translation of some of the documents from one language to another, or from earlier to current usage, may have resulted in distortion in meaning. Even contemporary writers in one's own language are frequently difficult to understand; many of them cannot be taken literally and can be understood only through knowledge of

their style of writing. And, of course, the problem is not made any easier when socio-economic and cultural differences—with local idioms and slang expressions—are introduced.

The establishment of the general reputation, integrity, and competence of its author, and the circumstances under which it was written, are of particular importance in determining the validity of the content of a historical document. At this point, internal and external criticisms are interdependent and complementary processes. For example, a document is less suspect if it can be established that it was written by Abraham Lincoln or George Washington on some topic with which they were familiar. If, on the other hand, the witness is suspected of bias, or considered a poor observer because of errors he is known to have made in similar connections, his testimony must be more severely questioned—and, of course, more readily rejected.

The historian must also attempt to appraise the motives of the writer. If an author is relatively unknown, it may be possible to appraise the general credibility of what he says on the basis of his position—for example, a government official or a minister of the church. In a sense, this is like the situation in a court of law, and the student might profit from developing the parallel. Thus, the testimony of a reputable person in the area of his competence is generally accepted, while a witness found lying, even in part, is frequently discredited *in toto*. Of course, the historian cannot accept or reject an entire document because one part is correct or erroneous; he must appraise each fact by itself. Questions, such as: Was the witness in a position to observe? Was he mentally competent? Did he stand to gain from his testimony or was he a disinterested observer? and so on, can help determine the acceptability of the testimony. On the other hand, the point can be carried too far; even the most disreputable witness may tell the truth occasionally, and a most adequate witness can be in error on occasions.

The circumstances under which a document was written are also important. For example, the statement by the then Senator Kennedy that 17 million Americans were going to bed hungry must be evaluated in the light of the fact that it was said in a campaign speech. Frequently people are in a position

in which they cannot be completely candid. Biographical writings are particularly suspect from the standpoint of accuracy, especially if the subject is still alive. In fact, in evaluating biographical documents, the relationship of the writer to the subject must be fully explored. If he is to place proper interpretation on his findings, the historian must analyze the motives of the writer. These may range all the way from monetary reward —though education is relatively free from this incentive—to friendship or enmity. It is generally accepted that the more disinterested the writer, the more likely he is to give a faithful reproduction of the facts.

Autobiographies are generally even more suspect. If the source is published after the death of the author, it is difficult to verify what a dead person has said, or even to know if what he recounted was something that he witnessed himself or something that he obtained secondhand. There is, for instance, the individual who in his writings takes credit for events in which he was involved, when in reality, he may have been a very minor operator in what occurred. Thus John Smith is generally considered such a braggard, that historians question very seriously his story of Pocahontas. Here again, the general credibility of the author is frequently an important clue to the validity of the content.

The validity of a historical "fact" can sometimes be verified by comparing it with the statements of other authors, though agreement may mean nothing more than that they have obtained their information from the same—perhaps erroneous—source. Thus, the common belief that the Declaration of Independence was signed on July 4, 1776 does not deny the fact that most of the signing was done on August 2, 1776. Similarly, one must be careful not to interpret the statement that American schools were desegregated as the result of the Supreme Court ruling of 1954 as meaning that racial segregation ended in America in 1954. When there is disagreement among authors, the historian must establish which, if any, is correct. This, he attempts to do on the basis of overall plausibility, reputation, independent corroboration, and general compatibility with other known facts.

The whole matter of numbers in historical writings is particularly bothersome. Not only are dates undependable, but it

is generally accepted that numbers used in documents, even as late as the Middle Ages, are so undependable that they are relatively meaningless. That Methuselah is alleged to have lived to the age of 969 means relatively little. The number 20 is used in Chaucer in the sense of many, and even today such expressions as 1001 cannot be taken literally. Similarly, statistics on enrollment, population, library holdings, and so on cannot always be accepted at face value. When there are conflicts between sources, one must depend on general credibility and give preference to the source with the greatest plausibility from the standpoint of internal consistency and agreement with other accepted facts.

Thus it is evident that historical research is an exacting task, calling for a high level of scholarship. Invariably, the historian will have to rely partially on sources that he can no longer verify; on many occasions he will have to rely on inferences based upon logical deduction in order to bridge gaps. At times, the historian will be unable to verify or to discredit the evidence before him, and yet he has to accept it or to reject it. In such cases, it probably is best for him to predicate his remarks with the phrase "according to . . . ," and thus safeguard his reputation and avoid the misuse of his status as a scientist in misleading his readers.

INTERPRETATION OF HISTORICAL DATA

Having established the authenticity and validity of his facts, the historian must address himself to the even more fundamental task of interpreting these facts in the light of his problem. In this, he must be especially aware of the limitations of his data. Because of the relative incompleteness and unverifiability of historical evidence, the interpretation of its significance requires the historian's greatest ingenuity and imagination. And yet, he must not let his imagination run away with him. This constitutes a major test of the historian's claim to scientific status.

Causation is a troublesome concept in science; it is doubly so in historical research where "causes" are in the nature of antecedents, or precipitating factors, rather than "causes" in the restricted scientific sense. Furthermore, historical causes are invariably complex, and a common error in interpretation is

oversimplification—for example, Caesar's ambition. Since history is actually a record of human behavior, causes in history are best considered on the basis of the motives of the participants. Thus, it might be more appropriate to consider the causes of World War II from the standpoint of the motives underlying the behavior of Hitler and his Nazi followers. However, behavior is based on a multiplicity of interacting motives of various vector strength, a fact which makes the task of the historian relatively difficult, though, of course, some insights into the general motivational structure of the participants can be obtained from the consistency and general pattern of their behavior. The problem involves the psychology of human behavior, and a historian, to be successful, should have some training in this area.

The historian must be very cautious in his use of analogy as a source of hypothesis or as frame of reference for interpretation. Because of the complexity of historical data, it is generally possible to draw parallels between one historical event and any number of others. Thus, the present administration can be compared to any one of the previous administrations from one standpoint or another, and historical parallels can be oriented in a number of directions depending on the historian's viewpoint. In the late thirties a parallel was often drawn between Spartan and Nazi education, for example. Any such comparison is invariably characterized as much by exceptions and differences in certain aspects, as it is by similarities in others, so that any attempt at extrapolation is at best risky.

In his interpretation of historical evidence, it is imperative that the historian not interpret data of different cultures or different historical periods from the standpoint of his personal standards—which, of course, do not apply. It is very difficult, for example, to put the brutality of the Middle Ages in proper focus.

The historian's goal, then, is not only to establish facts but also to determine trends which the data may suggest and generalizations which can be derived from the data. His task is one of synthesis and interpretation rather than mere summation. This calls for some frame of reference, and, of course, it must be recognized that historians differ in their interpretations of the same facts. Thus, World War II has been the sub-

ject of considerable disagreement with respect both to specific events and to their interpretation—a disagreement resulting from the superimposition of a difference in frame of reference on a relative lack of data rather than from a deliberate attempt at distortion.

THE WRITING OF THE REPORT

The writing of the historical report is a task involving the highest level of scholarship. While conducting and reporting research is never simple, it is doubly difficult in the field of history where so much depends on the ingenuity and the scholarliness of the investigator. Because of the relative lack of conclusive evidence on which valid generalizations can be established, it is generally accepted that the writing of historical research has to be a little more free in allowing a somewhat greater reliance on subjective interpretation of data. This does not, however, condone the distortion of the truth. Though the approach to the collection of data needs to be flexible, it must be sufficiently systematic to prevent unnecessary gaps or omissions.

The discontinuous and incomplete nature of historical evidence places a particular burden on the ingenuity and insight of the investigator in providing the required continuity. At the same time, it allows the historian plenty of room in which to show his scholarship in the insights which he displays into his subject, the plausibility and clarity of his interpretations, the ingenuity and creativity which he brings to the solution of his problem, and the adequacy of the writing of his report. Inasmuch as there is a need to interpret the data, as well as to record it, the historian must digest his material in order to attain historical perspective.

Of course, the discontinuity of historical evidence increases the danger of error, and if the gaps in the evidence are such that an attempt at interpretation is unsafe, the historian must be careful not to part company with his scholarship. He may either indicate that the gaps exist and stop there, or he may suggest a number of alternative solutions. He may even leave the task of presenting his data to the more reckless historical writer.

Although the historian is not permitted liberties with the truth, historical presentations need not be dull. The writer may

feel that to be historically accurate he must be cumbersome. Instead of stating simply that Columbus discovered America, he may want to point out that Columbus is alleged to have landed on an island now known as Watling Island. Gottschalk feels that though this may be commendable it probably is not justified. Since one can say even dull things in an interesting way, there is no justification for writers to get bogged down in battles and discoveries—in "the same slough of uninspired verbiage."[3] When the facts are of such a nature that they would have to be qualified repeatedly, and thus make the reading cumbersome, the solution might be to pass the task to a historical writer, who is not subject to the same restrictions as the historical researcher.

HISTORICAL RESEARCH—SCIENTIFIC?

A question that has been discussed repeatedly—though perhaps not profitably—is whether or not historical research is a scientific endeavor. The question is essentially academic; it is considered here for whatever light it sheds on the nature of the historical method. The whole issue centers around the definition of terms and the criteria used. If we accept the principle that science is oriented toward the discovery of laws capable of conclusive verification, historical research probably does not qualify as a science. As we have seen, historical research is characterized by a relative inability to establish control, by a complexity of relatively unverifiable and incomplete data, by a relative lack of acceptable criteria for the analysis of data, by a relative over-dependence on subjectivity of interpretation, and by the impossibility of empirical verification of its deductions. If, on the other hand, the criteria are defined on the basis of critical methods of discovery and of scholarship, then historical research frequently meets this requirement at the highest level.

There are three main tasks in historical research; 1. the collection of data; 2. the treatment and interpretation of data; and 3. the derivation of conclusions and generalizations. From a strict point of view, historical research can be criticized for failure to meet the criteria of science in all three tasks.

[3] Louis R. Gottschalk, *Understanding History: A Primer of Historical Method* (New York: Knopf, 1950), p. 10.

1. *The Collection of Data.* Historical data, as the basis for historical generalizations, are not comparable to the materials of the physical sciences. They have to be reconstructed, in many cases, from rather nebulous and essentially unverifiable sources. Historical facts are not "knowable" in the sense that the facts of the physical sciences are; they have to be inferred and accepted on the basis of plausibility. Also, though the scientific method attempts to arrive at a workable hypothesis on the basis of a comparison of a sufficiently large number of samples, historical research is generally based on unique events which occurred but once, and which cannot occur again.

2. *The Treatment and Interpretation of the Data.* The natural sciences are oriented toward experimentation. Historical problems, on the other hand, since they deal with unique events, cannot be experimented on and are verifiable only on the basis of logical deduction. As we have seen, it is very difficult for the historian to make an adequate analysis of his data. Often he must deal with material that has to be deciphered or translated. Interpretation of the present is difficult, especially when such things as satire, allusions, metaphors and other figurative liberties are involved. It is even more difficult to treat a different period of history and/or a different national or cultural group—which automatically introduces such complications as the difficulty of translation, differences in the use of words, and differences in customs and mental outlook.

3. *The Products of Historical Research.* Historical research can also be criticized from the standpoint of the products it is supposed to provide. Basically, research is oriented to derivation of laws and principles expressing certain regularities among phenomena. This borders on the concept of causation, which is especially confusing in the case of historical events. Thus, the assassination of Archduke Ferdinand can be considered the precipitating "cause" of World War I—just as perhaps marriage may be considered the only readily identifiable "cause" of divorce—but neither statement tells the whole story. Similarly, as Bertrand Russell[4] points out, Eli Whitney can be considered the "cause" of the War between the States, since his invention of the cotton gin led to a renewed interest in slavery.

[4] Bertrand Russell, *The Impact of Science on Society* (New York: Simon and Schuster, 1953), pp. 20–21.

Thus, while in the physical sciences the goal of the researcher is the derivation of verifiable conclusions that can eventually become laws, generalizations of historical evidence, since they are based on unique events that technically occurred only once and can never occur again, are somewhat meaningless. In fact, the concept of historical "laws" is perhaps self-contradictory. Of course, this point of view can be challenged since the distinction between physical and historical laws is essentially a matter of degree rather than of kind. Certain laws of a historical nature—for example, the law of supply and demand, the law of diminishing returns, and many others —possess the same basic properties as other scientific laws.

In summary, historical research can be considered as lacking a number of the characteristics of the scientific method, interpreted in its narrow sense. For that matter, many aspects of educational and sociological research today do not meet the strict requirements of science as they are defined in the physical sciences. On the other hand, a number of historical facts have been established beyond reasonable doubt; it is accepted that Christopher Columbus discovered America, that the Pilgrim Fathers landed at Plymouth Rock on December 21, 1620, that the Chinese invented gunpowder, that certain documents are frauds, and that Pittsburgh won the World Series in 1960. In fact, as pointed out by Gottschalk,[5] the amazing thing is not that historians disagree but that they agree as much as they do. Indeed, the ingenuity with which historians have proceeded in such discoveries as Champollion's deciphering of the Rosetta Stone, which provided the key to Egyptian hieroglyphics, as well as their systematic and painstaking approach to such significant problems as the authorship of the Shakespearian plays or the existence of Moses, reflect a fascinating degree of scholarship. It does not make sense to reject historical research as unscientific, and then, simply because they have been subjected to the legerdemain of statistical treatment, to brand as scientific questionnaire studies, with their usual inadequacy in the areas of non-return, misinterpretation, and other inherent weaknesses.

Historians do find a common ground with other scientists in the scholarly nature of their efforts to seek truth within the framework of the data with which they have to deal. Historical

[5] Gottschalk, *op. cit.*, p. 1.

research must adhere to the same principles and practices, and the same scholarship and accuracy, which characterize all scientific research. It must follow the same steps of the identification, selection, and delimitation of the problem; the formulation of hypotheses; the collection, organization, and verification of data; and the testing of the hypotheses. More specifically, the historian as a scientist must display a complete mastery of the material with which he deals. He must display originality, ingenuity, creativity, and critical insight into the meaning of facts, and he must maintain the usual scientific objectivity, for, though many gaps in the data will have to be filled according to his best judgment, he is still bound by the rules of science.

At all times, the historian must operate inductively—that is, rather than starting with a hypothesis and then marshalling the facts to support it, he must rely on deduction only to check the plausibility of his hypothesis or tentative generalization. In connection with the hypothesis that certain plays were written by a young playwright from Stratford named William Shakespeare, for example, one might reason deductively that in order to have been the author Shakespeare would have had to be rather well educated. Such deductive reasoning can, of course, lead to the rejection of certain hypotheses, and thus orient the investigation toward more fruitful leads. Finally, the historical report must meet the usual standards of scientific and scholarly writing.

In view of the difficulties inherent in historical research, the student must be particularly careful in selecting a historical topic for his thesis or dissertation. He must realize that it is difficult to obtain historical evidence of an acceptable scientific nature. The major problem is, of course, the relative unavailability of historical evidence of acceptable validity on the basis of which gaps in knowledge can be bridged, contrary and conflicting evidence can be reconciled, and valid generalizations reached. It is necessary to check the dependability of one's sources, the validity of the data, and, despite the many limitations these may involve, the investigator must still reach dependable generalizations.

Historical research has fallen into some degree of disrepute because of its excessive reliance on subjectivity and secondary sources of dubious value, and the choice of an historical topic

as a doctoral or master's thesis has, in general, been discouraged. Actually, of course, like philosophical studies, historical studies can provide a perspective for many educational problems in relation to which we must constantly make important decisions, and an adequate historical study can undoubtedly make a major contribution to the cause of education.

CRITERIA OF HISTORICAL RESEARCH

A number of criteria on the basis of which historical research may be evaluated can be obtained readily from the preceding discussion. A few of the major points are included in the following checklist.

1. PROBLEM. Has the problem been defined clearly? It is difficult enough to conduct historical research adequately without adding to the confusion by starting out with a nebulous problem. Is the problem capable of solution? Is it within the competence of the investigator?
2. DATA. Are data of a primary nature available in sufficient completeness to provide a solution, or has there been an over-dependence on secondary or unverifiable sources?
3. ANALYSIS. Has the dependability of the data been adequately established? Has the relevance of the data been adequately explored?
4. INTERPRETATION. Does the author display adequate mastery of his data and insight into their relative significance? Does he display adequate historical perspective? Does he maintain his objectivity? Are his hypotheses plausible? Have they been adequately tested? Does he see the relationship between his data and other "historical facts"?
5. PRESENTATION. Does the style of writing attract as well as inform? Does the report make a contribution on the basis of newly discovered data or new interpretations, or is it simply "uninspired hackwork"? Does it reflect scholarliness?

POSSIBLE RESEARCH AREAS

The number of historical studies that could be conducted with profit to education is relatively unlimited. These range from those that are primarily of local interest to those that have rather widespread appeal. They might cover any aspect of educational practice—curriculum, methods of instruction, school organization, and so on—at any period of its evo-

lution from the days of the Greeks (and even earlier) to the present. All have a history, an understanding of which can be of considerable value in giving present practice perspective and orientation. The following are among the broad general areas from which specific topics for investigation might be selected.

1. The development of current movements in American education. Some of the more significant movements of the past—Herbartianism, the Montesorri and Winnetka systems, and progressive education—have received considerable attention from educational historians. Similar studies could be made of "honors" programs and other innovations and special programs designed to provide more effective education, of pupil guidance, of the city-college or private-school movement, of racial integration, and of the evolution of such perpetual problems as the report card. The gradual changes toward greater functionalism taking place in school buildings and school furniture could also be traced profitably. Even the "history" of the lay criticism of public education over the centuries could provide valuable insights into the role of the school as an agency of society.

   Investigation affecting teachers might be of even greater interest. Possible areas are: the fifth-year program, screening in teacher-education institutions, merit pay, the use of strikes as a bargaining tool, the social status of the teacher in the community, teacher aides, teacher certification, and so on.
2. The evolution of current practices in classroom organization (the graded or ungraded school, team teaching); instructional procedures (the problem-solving approach, the integrated use of the library or of audio-visual aids); the curriculum (changes in the approaches to mathematics, science, and foreign languages); pupil personnel (changing emphasis in discipline, mental health, moral education, and the co-curricular program); and so on. The evolution of the philosophy underlying educational practice—for example, the "pupil-activity" concept of learning—might also be of interest.
3. The contributions of leading educators and their influence on current educational practice and thought, of leading universities, and of important professional organizations. Studies could also be made of the evolution of special agencies and offices on the American educational scene—for example,

the state superintendency, the subject area supervisor, or even the school of education.

4. Problems of special interest: the history of Indian or Negro mission schools or even of a local college, perhaps on the occasion of an anniversary.

A number of studies typical of the areas listed are readily available in the literature. Many are classics in their field; others, however, are lacking in documentation, perhaps as a result of the unavailability of more adequate sources. Some are rather dated and, in some instances, could be brought up to date. The following are among the better known:

Cheyney, Edward P. *History of the University of Pennslyvania, 1740–1940*. Philadelphia: University of Pennsylvania Press, 1940.

Clifton, John L. *Ten Famous American Educators*. Columbus: Adams, 1933.

Coon, Horace. *Columbia: Colossus on the Hudson*. New York: Dutton, 1947.

Graves, Frank P. *Great Educators of Three Centuries*. New York: Macmillan, 1912.

Pangburn, Jessie M. *The Evolution of the American Teachers College*. T. C. Contributions to Education, No. 500. New York: Teachers College, Columbia University, 1932.

Ryan, W. Carson. *Studies in Early Graduate Education*. New York: Carnegie Foundation for the Advancement of Teaching, 1939.

Sears, Jessie B. *Cubberley of Stanford: and His Contributions to American Education*. Stanford: Stanford University Press, 1957.

Woody, Thomas. *A History of the Education of Women in the United States*. 2 vols.; Lancaster: Science Press, 1929.

CLASSICAL STUDIES IN HISTORICAL RESEARCH

The most significant discovery of historical data in recent years is, of course, that of the Dead Sea Scrolls which have been confirmed as genuine documents left by Jewish tribes at the (approximate) time of Christ. The first scrolls were discovered in 1947, but they did not become usable until 1956 when a process of spraying them with glue and baking them, so that they could be sawed open and photographed without disintegrating, was discovered. The Scrolls are now in process of being

translated, and their full significance, as well as the validity of their contents, is yet to be determined, but they will undoubtedly be of primary importance in the understanding of the Jews of that particular period of history.

Champollion's deciphering of the Rosetta Stone (1822), which opened the whole area of Egyptian hieroglyphics, has already been mentioned. Of importance in our own country's history is the Kensington Stone found in Minnesota in 1898. Considered a fraud for some years—and, at one time, used as a doorstop—it is now accepted as valid evidence of the presence of Norwegian nationals in the middle United States in an early period in our history.

Although no such spectacular historical "discoveries" are to be found in the field of education, special mention must be made of Cubberley's *Public Education in the United States*,[6] which gives a comprehensive coverage of the various movements in American education with their sociological and philosophical significance. Graves[7] gives a correspondingly adequate coverage of world education from the early period to the present. A more recent publication (1961) is Cremin's *Transformation of the School, 1876–1957*.[8]

DOCUMENTARY RESEARCH

Very closely related to historical research is documentary research—that is, research based on documents and records. Though the distinction is not always clear-cut, documentary research differs from historical research in that it usually excludes remains as a source of evidence, and, conversely, may include the study of contemporary documents, such as might be involved in deciphering enemy codes. On the other hand, this distinction is not always binding; on occasion documentary research has concerned itself with utensils, pottery, and even natural specimens, such as rocks and fossils.

The location of documents is often a chance affair. Many

[6] Ellwood P. Cubberley, *Public Education in the United States* (Boston: Houghton-Mifflin, 1947).

[7] Albert D. Graves, *A History of Education:* 1. *Before the Middle Ages*, 2. *During the Middle Ages*, 3. *In Modern Times*. (New York: Macmillan, 1909, 1910, and 1913).

[8] Laurence A. Cremin, *Transformation of the School, 1876–1957* (New York: Knopf, 1961).

stories are told of famous letters and other documents retrieved from attics or junk dealers and, not infrequently, from the edge of the furnace. More typically, it involves a great expenditure of time, energy, and effort, as well as ingenuity, in tracing one lead after another until documents are located and, frequently, a great deal of persuasion before they are obtained for study.

The crucial aspects of documentary research, like those of historical research, are validating the data and interpreting their significance. Legal documents tend to be very dependable, but ordinary records frequently are in considerable error. Statistical data are rarely comparable; in devising an index of business conditions, for instance, one frequently finds sizable discrepancies over the years in such things as whether office workers in industrial firms are included among *industrial workers,* whether sales data are adjusted for seasonal variations, and so on. College enrollment figures or library holdings are rarely comparable from school to school. The problem is even greater when the data are obtained from different documentary sources. The federal government has established considerable uniformity in the data it reports, but there is no such uniformity in local data or in data collected by various industrial or commercial agencies. The problem becomes even more impossible when foreign nations are involved. Thus, infant mortality in certain undeveloped nations may be abnormally low simply because birth records are extremely incomplete; many infants who die early in life never become part of either the birth or the death records.

BIBLIOGRAPHIC RESEARCH

Bibliographical research is oriented toward the integration and critical synthesis of the status of a given problem. In a sense, therefore, it resembles a term paper except that it is more critical and of a higher level of worthwhileness, comprehensiveness, and complexity, and generally is frowned on as a doctoral dissertation. On the other hand, such a study, made by a person with considerable insight into the overall problem, can frequently make a significant contribution to education by structuring the field and identifying the areas in need of further investigation. A great deal can be gained, for example, by hav-

ing someone clarify such issues as motivation or the differences in viewpoint among the various schools of psychology or of philosophy. Bibliographic research deserves a better status than it has had. However, bibliographic research is difficult to conduct adequately, particularly by a graduate student, who is not likely to have the degree of insight necessary to do such a study justice. For that reason, the general reluctance of graduate faculties to accept bibliographic studies in fulfillment of the research requirements for the degree is probably justified.

LEGAL RESEARCH

In view of the legal responsibilities connected with the various aspects of managing the school, legal research is of particular interest to school administrators, but it also involves, in various degrees of directness, every member of the profession. Is the chemistry teacher responsible for accidents occurring in his chemistry class? Is the football coach responsible for an injury to a player? Can a teacher detain a student so long that he will miss his bus? These are some of the questions which require answers, and, though answers are available, they are generally complex and involve a number of provisions, special considerations, and technicalities.

Legal research is subject to the same general requirements as are other forms of research. In nature, it most closely resembles bibliographical research. The task is to find and summarize pertinent statutes, to trace further legal developments through related court decisions, and finally to analyze the decisions in the light of the problem being investigated. The last step is the writing of the report which must convey legal information to educators and laymen who are not themselves legally trained. Obviously legal research calls for special training in the field of law, and anyone without this training is not competent to do this type of research. In fact, in view of their complexity, such studies are generally best undertaken by such organizations as the National Education Association, rather than by a graduate student working toward a degree.

SUMMARY

1. Because of the difficulty of obtaining dependable data, historical research is among the most difficult to conduct adequately,

and the student should exercise caution in selecting a historical problem for the fulfillment of the research requirement for his degree. On the other hand, a survey of the past can frequently provide valuable insights into present practices, and education might profit from a de-emphasis of its present reluctance to sponsor historical research for thesis or dissertation purposes.

2. The historian generally conceives his task to be the interpretation of the past in the light of a certain point of reference, rather than simply the development of a chronicle of events. The three major points of view from which historical perspective is superimposed on historical data are the scientific, the biological, and the humanistic. Among the more specific orientations are the technological, the geographic, the great-man, and the eclectic theories.

3. Historical evidence is almost invariably inadequate: not only are historical events unique and incapable of verification through duplication, but records are invariably lacking from the standpoint of accuracy, completeness, impartiality, and so on. This is particularly evident in some types of data, for example, dates and numbers.

4. The historian must rely on primary sources for the bulk of his information and where such gaps exist in available primary sources that he has to place undue reliance on secondary sources and/or his imagination to bridge the gaps, he should probably refrain from undertaking the study. It must, of course, be realized that, while secondary sources are frequently undependable, they are sometimes on the contrary most trustworthy. The historian uses all the evidence at his disposal, but he must take special care to ensure its validity by subjecting it to rigorous test.

5. Historical evidence must be carefully evaluated from the standpoint of both its authenticity and its validity, and very elaborate techniques have been devised to preclude the perpetration of frauds. On the other hand, while such methods of detection can lead to the rejection of historical evidence as false or fraudulent, they can lead to its acceptance only on the basis of plausibility.

6. The accumulation and validation of historical data, while crucial, is only a step to the even more important task of interpreting their significance. Here the historian is on extremely subjective grounds and he must be careful not to part company with his scholarship. The establishment of causation is particularly precarious, for example. On the other hand it is precisely through the display of his grasp of the field, the clarity and plausibility of his interpretations, his ability to bridge gaps, the continuity and the perspective which he superimposes on these data to make them meaningful that the historian establishes his claim to scientific status.

7. While the writing of the historical report must unavoidably—

and desirably—allow for a somewhat greater degree of freedom in the use of subjectivity than does the usual research report, this is not a license for the historian to let his imagination and his personal biases distort the facts.

8. Whether historical research qualifies as a scientific endeavor depends on the criteria used. While historical research cannot meet some of the tests of the scientific method, interpreted in the narrow sense of its use in the physical sciences, it does qualify from the standpoint of its subscription to the same principles and the same general scholarship and accuracy which characterizes all scientific research.

9. Documentary, bibliographic, and legal research, though not strictly "historical" in nature, share somewhat the same problems with historical research, particularly from the standpoint of the incompleteness, the discontinuity, and unverifiability of the data and the crucial role which the investigator's insight plays in the interpretation of their significance.

PROJECTS and QUESTIONS

1. Make a historical study of the development of historical research. Appraise its present status and its current trends.
2. Make a documentary study of the present status of educational research as revealed by the professional literature (including textbooks).
3. Identify common points of agreement among the educational leaders whose influence is incorporated in present educational practice—for example, Rousseau, Herbart, Dewey—with respect to such issues as the relative role of the teacher in the learning of the child.
4. Trace the evolution of certain basic concepts underlying educational theory and practice—for example, the concept of pupil activity as a factor in the effectiveness of his learning.
5. Evaluate the biography of a great scientist from the standpoint of its compliance with the criteria of a scientific document.

SELECTED REFERENCES

AUSUBEL, HERMAN. *Historians and Their Crafts: A Study of the Presidential Addresses of the American Historical Association, 1884–1945*. New York: Columbia University Press, 1950.

BAUER, RICHARD H. "Bibliographic Guide to the Study of History: With Selected References to Historical Research," *Social Studies*, 43 (February 1952):64–9.

BENJAMIN, HAROLD H. "An Approach to the Study of Causation in Educational History," *History of Education Journal*, 6 (Fall 1954):137–52.

BORROWMAN, MERLE L. "History of Education," in Harris, Chester W.

(ed.), *Enclyclopedia of Educational Research.* 3rd ed.; New York: Macmillan, 1960.

BRICKMAN, WILLIAM W. *A Guide to Research in Educational History.* New York: New York University Bookstore, 1949.

BROOKS, FOWLER D. (ed.). *The Use of Background in the Interpretation of Education Issues.* 25th Yearbook, National Society of Colleges of Teacher Education. Chicago: University of Chicago Press, 1937.

BRUBACHER, JOHN S. *A History of the Problems of Education.* New York: McGraw-Hill, 1947.

CLIFTON, JOHN L. *Ten Famous American Educators.* Columbus: Adams, 1933.

CREMIN, LAURENCE A. *Transformation of the School, 1876–1957.* New York: Knopf, 1961.

CUBBERLY, ELLWOOD P. *Public Education in the United States.* Boston: Houghton-Mifflin, 1947.

DAVIES, A. POWELL. *The Meaning of the Dead Sea Scrolls.* London: Muller, 1957.

DUNLAP, KNIGHT. "The Historical Method in Psychology," *Journal of General Psychology,* 24 (January 1941): 49–62.

GARRAGHAN, GILBERT J. *A Guide to the Historical Method.* New York: Fordham University Press, 1946.

GOOD, HARRY G. "Historical Research in Education," *Educational Research Bulletin,* 9 (February 5, 1930): 7–18, 39–47, 74–8.

———. "The Possibilities of Historical Research," *Journal of Educational Research,* 29 (October 1939): 148–53.

GRAVES, FRANK P. *A History of Education:* (a) *Before the Middle Ages* (1909); (b) *During the Middle Ages* (1910); and (c) *In Modern Times* (1913). New York: Macmillan.

———. *Great Educators of Three Centuries.* New York: Macmillan, 1912.

GOTTSCHALK, LOUIS R. *Understanding History: A Primer of the Historical Method.* New York: Knopf, 1950.

HOCKETT, HOMER C. *Introduction to Research in American History.* New York: Macmillan, 1948.

———. *The Critical Method in Historical Research and Writing.* New York: Macmillan, 1955.

JOHNSON, LEIGHTON H. "Education Needs Historical Studies," *Phi Delta Kappan,* 36 (January 1955): 157–9.

PANGBURG, JESSIE M. *The Evolution of the American Teachers College.* T. C. Contributions to Education, No. 500. New York: Teachers College, Columbia University, 1932.

PERDEW, PHILIP W. "History of Education and the Educational Professions," *Educational Forum,* 12 (March 1948): 311–23.

———. "Criteria of Research in Educational History," *Journal of Educational Research,* 44 (November 1950): 217–23.

PERDEW, PHILIP W. "Analysis of Research in Educational History," *Phi Delta Kappan,* 32 (December 1950) : 134–6.

PETERSON, HOUSTON (ed.). *Great Teachers.* New Brunswick: Rutgers University Press, 1946.

QUIRKE, ARTHUR J. *Forged, Anonymous, and Suspect Documents.* London: Routledge, 1930.

RENIER, GUSTAF J. *History: Its Purpose and Method.* Boston: Beacon Press, 1950.

RYAN, W. CARSON. *Studies in Early Graduate Education.* New York: Carnegie Foundation for the Advancement of Teaching, 1939.

SCHINDLER, MARGARET C. "Fictitious Biography," *American Historical Review,* 42 (July 1937) : 680–9.

SEARS, JESSIE B. *Cubberley of Stanford and his Contribution to American Education.* Stanford: Stanford University Press, 1957.

STRAUS, WILLIAM L. "The Great Piltdown Hoax," *Science,* 119 (February 26, 1954) : 265–9.

WILLIAMS, E. I. F. *Horace Mann: Educational Statesman.* New York: Macmillan, 1937.

WILSON, EDMUND. *The Scrolls from the Dead Sea.* New York: Oxford University Press, 1955.

WOODY, THOMAS. *A History of the Education of Women in the United States.* 2 vols. Lancaster: Science Press, 1929.

———. "Of History and Its Methods," *Journal of Experimental Education,* 15 (March 1947) : 175–201.

In intensity of feeling, and not in statistics, lies the power to move the world. But by statistics must this power be guided if it would move the world aright.

CHARLES BOOTH

10 The Survey: Descriptive Studies

No category of educational research is more widely used than the type known variously as the *survey*, the *normative-survey*, *status* and *descriptive research*. This is a broad classification comprising a variety of specific techniques and procedures, all similar from the standpoint of purpose—that is, to establish the status of the phenomenon under investigation.

NATURE OF SURVEY RESEARCH

Although it is not possible to make clear-cut distinctions between these studies and the other research classifications, general differences can be pointed out. A fairly clear line can be drawn between survey studies and historical studies on the basis of time: the latter deals with the past, the former with the present.[1] Surveys differ from experimental studies in purpose: Surveys are oriented toward the determination of the status of

[1] Less clear is the distinction between surveys and documentary and legal research whose primary purpose is to "survey" existing documents. As previously noted, both documentary and legal research could have been included in the present rather than in the previous chapter.

a given phenomenon rather than toward the isolation of causative factors. Survey studies differ from case studies in that surveys are generally based on large cross-sectional samples, while case studies are oriented to the more intensive and longitudinal study of a smaller sample and, like experimentation, attempt to isolate antecedents or causes of the phenomenon under investigation.

The comparison between survey studies and other forms of educational research is complicated by a number of subsidiary outcomes which often accrue as by-products of surveys. For instance, the comparison of the status of two or more groups subjected to differential treatment approximates the experimental method.[2] Similarly, successive surveys can establish trends and permit the prediction of the likely status of phenomena. Census figures from one census to the next, for example, are a valuable gauge of national growth. Furthermore, when a distinct break in a given trend can be associated with a procedural change or with the introduction of a certain factor at that point, the break can be considered a crude experiment on the relative effect of the change. While subsidiary results are frequently of major importance, they are merely by-products. The primary goal of the survey is the investigation of the present status of phenomena.

While surveys are, on the whole, relatively less scientifically sophisticated than most other research techniques, they vary in complexity and sophistication. At one extreme, they constitute nothing more than a clerical fact-finding approach to the study of local problems conducted on a one-shot basis, without any significant research purpose—for example, a survey of the academic qualifications of school superintendents. At the other extreme are surveys that bear directly on significant interrelationships among phenomena. Surveys of the reactions of inmates of concentration camps, for example, have provided definite insight in the psychology of the human personality under conditions of psychological stress. Terman's studies of the gifted[3] have likewise been of considerable practical and theoretical significance

[2] It must be realized that such natural experiments are often crude and lacking in adequate control.

[3] See Chapter 15.

Historically, surveys date to the first census ordered by Caesar Augustus. They vary in subject from such topics as the duties and responsibilities of the school superintendent, the activities of the classroom teacher to the attitudes of school personnel or the public on a wide variety of educational issues. In scope, survey studies range from such vast undertakings as the decennial census of the Bureau of the Census to the on-the-spur-of-the-moment poll of the smoking preferences of college students.

Purpose of Surveys

Educational surveys are particularly versatile and practical, especially for the administrator, in that they identify present conditions and point to present needs. They cannot make the decisions for the administrator, but they can provide him with information on which to base sound decisions. Surveys are so obviously useful, in fact, that administrators tend to rely on them too exclusively, and to base crucial decisions on a survey of opinions—often poorly sampled. Surveys are of the present and, if used simply for the purpose of seeing what has been attained to date, are relatively useless. On the other hand, by providing the basis for decisions for improvement, they can be decidedly practical.

Surveys must do more than merely uncover data; they must interpret, synthesize, and integrate these data and point to implications and interrelationships. And, while the fact-finding aspects of the survey are occasionally semi-clerical in nature, there is ample opportunity for the investigator to display ingenuity and scholarliness in his interpretation of the data and in his understanding of their strengths and weaknesses, their interrelationships, their apparent antecedents, and, especially, their implications.

Survey research, like all other research, must begin with a definite problem and be oriented toward the eventual derivation of valid generalizations. The survey makes its maximum contribution when it originates from a problem existing within the framework of theory, and when it is oriented toward the identification of factors and relationships worthy of investigation under more rigorously controlled conditions. Since it is rarely possible to achieve control of extraneous factors within

the setting of the natural situation, the survey is not generally capable of testing specific hypotheses. As a method of research, it represents a step of intermediate scientific sophistication by which semi-crude relationships among phenomena are explored. It is a scientific technique only insofar as it strives for all the precision of which it is capable.

The survey constitutes a primitive type of research in that the investigation of any problem must begin with a "survey" of its nature before it can move into the more structured and rigorous phases. At its most elementary stage, the survey is concerned with determining the immediate status of a given phenomenon. More important from the standpoint of its role as a technique in the development of educational science, however, is the extension of this clarification of the problem into the development of further insights and, eventually, into the derivation of hypotheses to be incorporated into more adequate investigations at the experimental level. Thus, its purpose is both immediate and long-range.

The survey is more realistic than the experiment in that it investigates phenomena in their natural setting. This is, of course, a great strength in the early stages of the investigation of a problem in that it affords flexibility and versatility. In the latter stages of investigation, however, this strength becomes a weakness because the lack of control precludes a definitive test of crucial hypotheses. Unfortunately, though the survey should be a steppingstone to more precise investigations, in practice this second step is frequently overlooked. Too often surveys are made of problems that lead nowhere, that have no significant purpose or that are oriented toward meaningless topics. On the other hand, it does not follow that the survey is an inferior type of research; the concept of inferiority does not belong here since the answer that is needed depends on the type of question that is raised, and, certainly with respect to certain types of problems—for example, the attitudes of children toward cheating—the answer must be derived from a survey rather than from some more sophisticated approach.

Classification of Surveys

Survey studies can be divided into any number of subcategories, depending on the basis and purpose of classification.

Probably the most basic breakdown is to separate them into *descriptive* studies, which are oriented toward the description of the present status of a given phenomenon, and *analytical* studies, in which phenomena are analyzed according to their basic components. Along a different continuum, survey studies can also be classified according to the instruments and techniques used—for example, questionnaire, interview, observation, and so on. Although neither of these breakdowns is clear-cut, this dual system of classification seems to have merit from an operational, as well as from an organizational point of view, and will be used as the basis of the present discussion. This chapter will be devoted to descriptive studies.

Special Problems

Two problems which are of importance in all research are particularly crucial in surveys.

1. The problem of sampling is of primary concern in all survey studies, for unless the sample on the basis of which the data are collected is representative of the population selected for investigation, the conclusions drawn cannot apply to that population.

2. The validity of the instruments or techniques used in gathering the data is crucial to the validity of the conclusions that are derived from surveys. To the extent that the instruments used are not valid—and one must remember that validity applies to a particular situation under specific conditions—the results obtained cannot be interpreted nor can generalizations be reached.[4]

SURVEY TESTING

Undoubtedly, the most systematic survey research conducted in our public schools is the standardized academic achievement testing program. Every year, school districts spend thousands of dollars to appraise the outcome of their teaching efforts. In addition, there is the somewhat less comprehensive, but nonetheless highly organized, program of pupil

[4] As we have seen in Chapter 4, reliability, on the other hand, is of minor importance. Although unreliability is a serious problem in the interpretation of individual scores, research is concerned with group values, which—provided adequate samples are used—are relatively unaffected by errors of unreliability since, by their very nature, they are self-cancelling.

appraisal in intelligence, special aptitude, personality adjustment, and vocational interest.

A distinction needs to be made between the guidance function and the research function of such testing, since, though the two are not independent, we are concerned here only with the latter. Research is interested in groups—that is, it attempts to derive generalizations which are applicable beyond the individual case. Survey testing, as a research activity, usually is interested in comparing the achievement of the group—a class, a school, or a system—with the group on which the test was standardized.[5] In contrast, the guidance approach is interested in the child as an individual.

Problems in Survey Testing

Of the two major problems connected with survey research, sampling generally is of minor importance in survey testing, since in most school situations the total population in the grades concerned is tested. Any attempts to use volunteers or individuals selected on a judgmental basis to represent the school would, of course, be taboo. The problem of the validity of the instruments used, on the other hand, is both crucial and difficult to handle adequately. The major difficulty is the applicability of the test norms to the particular group under study. It must first be realized that many instruments have been inadequately standardized. But even more detrimental to meaningful results is the question of the applicability of the norms of a given test to a particular group. If we accept the principle that any test score is valid to the extent, and only to the extent, that the background of the testee is comparable to that of the group on which the test was standardized, how does one interpret the performance of children in a rural community on a test standardized on city children? Or the performance of children on a test in arithmetic which covers decimals when, because of the arrangement of the local curriculum, the unit of decimals is postponed one grade? Just how much below national norms is it permissible for a school in the slums to be? All of these questions point to the inescapable fact that the

[5] Inter-class or inter-school comparisons are sometimes made. From a research point of view, such comparisons are to be condemned, since the control necessary to make such comparisons meaningful is frequently lacking.

measurement of performance is simply a step to the more important task of interpreting that performance with respect to the objectives which the school feels it can legitimately attain.

Test-wisedness on the part of the student is another important factor to consider in the interpretation of the results of testing. This factor would be particularly invalidating when teachers prepare their students for the administration of the survey instrument by reviewing with an equivalent form of the same test. Also worth mentioning is the all-too-frequent practice of invalidating the test norms by "teaching toward the test." Inasmuch as the norm group did not have the benefit of this orientation, any comparison of the performance of the practiced class and the norms is completely meaningless.

Uses of Survey Testing Results

The benefits to be derived from periodic appraisals of the work of the school are undoubtedly great. Such studies not only can point to gaps and weaknesses in the program but also can serve to keep the whole system alert.[6] At the college level, entrance examinations enable the school to appraise its functions in relation to the students whom it undertakes to serve. Coupled with analyses of student grades, studies of admission-test performance help keep the school, and the units within the school, on an even keel and can be useful in such policy decisions as calibrating grading policy to the level of the students admitted.

If testing studies are to be of benefit to the school system and to the children, however, they have to be carried out correctly—or not at all. Since policy decisions are generally no better than the data on which they are based, and since unwise decisions can cause considerable harm, there appears to be little room for incompetence here. In any school system, the appointment of a director of testing who has considerable background both in the principles of measurements and in research methods, working through relatively adequate test chairman in each of the schools, seems to be essential if we are to justify such a program.

[6] On the other hand, appraised from a philosophical and pedagogical point of view, overemphasis on such evaluation can negate the very things for which the school exists; from a research point of view, it can invalidate any comparison with test norms.

THE QUESTIONNAIRE

The Questionnaire As a Research Tool

Probably no instrument of research has been more subject to censure than the questionnaire. Yet it continues to be the most used—and the most abused—instrument in educational research as both graduate students and professional agencies continue to rely on it. The questionnaire apparently dates back to Horace Mann, who is credited with having used it as a research tool in 1847. Its abuse—both in quantity and in lack of quality—reached such proportions in the post-World War I period that the National Education Association in 1930 devoted one of its most extensive articles to the consideration of the problem. It was noted, for instance, that some school superintendents received as many as one hundred questionnaires per year, most of them of a very inadequate nature. Despite its recognition of flagrant abuses, the N.E.A. study concluded that its findings did not support blanket condemnation of the questionnaire as an instrument of research, but rather pointed to an immediate need for its drastic improvement. The N.E.A. made specific recommendations for dealing with the problem; it even provided an informal rating scale for evaluating the questionnaires received, and advised its members to reply to a questionnaire only when its quality from the standpoint of sponsorship, worthiness of the topic, organization, and so on merited such reply.

As a result of such resistance, there has been a decline in the use of the questionnaire as a research instrument, together with a clarification of its proper use and an improvement in its quality. Today its weaknesses and limitations—as well as its strengths—are more clearly recognized, and a more serious attempt is made to limit its use to situations where it is appropriate. It is recognized that its weaknesses are not insurmountable. The problem is one of deciding when it is appropriate to use it—for instance, in preference to the interview or the experiment—and then of ensuring that it meets acceptable levels of adequacy. In other words, the questionnaire has definite advantages which must be weighed against its disadvantages, and its validity must be considered in the specific case.

The first problem to be faced in planning a questionnaire survey is, obviously, to decide whether an adequate answer can be obtained by a survey, or whether recourse should be made to more precise techniques. This calls for an understanding of the relative advantages both of surveys and of other forms of research. Too frequently a survey is made when a valid answer can come only from experimentation. Thus, a teacher might attempt to solve the problem of whether or not training in phonics promotes greater proficiency in reading by surveying the opinions of other teachers, who are equally ignorant of the answer. The decision must be based on a clear conception of specifically what the investigator wants to determine, and the kind of data necessary to answer the questions which the problem entails.

Assuming that a survey is indicated, the investigator needs to determine whether the questionnaire is the most adequate source of survey information in this particular case. This choice is made on the basis of the relative advantages and disadvantages of each of the relevant survey techniques in relation to the problem and the situation involved—or in technical terms, it is necessary to choose the best instrument from the standpoint of validity, reliability, usability.

Advantages and Disadvantages of the Questionnaire

The discussion of the relative advantages of the questionnaire must be restricted to what constitutes a relevant comparison. There is nothing particularly enlightening about weighing the merits of the questionnaire against those of the experiment, for example, since they are designed for essentially different purposes.

The choice of the questionnaire in preference to other survey techniques is generally a matter of weighing its strengths and weaknesses against those of the interview, with which it is most nearly interchangeable. In fact, some authors insist that the term *mailed questionnaire* be used to distinguish between the questionnaire that is mailed and that used as a guide in interviewing. The discussion will be oriented, therefore, toward a comparison of these two techniques on the basis of the usual criteria of validity, reliability and usability.

Among the major advantages of the questionnaire is that

it permits wide coverage for a minimum expense both in money and effort. It affords not only wider geographic coverage than any other technique, but it also reaches persons who are difficult to contact. This greater coverage makes for greater validity in the results through promoting the selection of a larger and more representative sample.

Particularly when it does not call for a signature or other means of identification, the questionnaire may, because of its greater impersonality, elicit more candid and more objective replies. Thus, depending on the topic—for example, the reactions of students toward their school—it may draw more valid responses though, to be sure, a skillfully conducted interview can frequently obtain equally good results. On the other hand, the questionnaire does not permit the investigator to note the apparent reluctance or evasiveness of his respondent, a matter which is better handled through the interview, nor does it permit the investigator to follow through on misunderstood questions or evasive answers.

The questionnaire also permits more considered answers. In an interview if the respondent does not have the information, he may still give an answer rather than admit his ignorance. The questionnaire is more adequate in situations in which the respondent has to check his information. The use of the questionnaire is also indicated in situations in which group consultation would result in more valid information. The questionnaire allows greater uniformity in the manner in which the questions are posed, and thus ensures greater comparability in the answers. This does not, of course, ensure truth, and at times a more valid answer may be obtained by phrasing the questions differentially in order to communicate more effectively with persons of the different sub-classifications within the population surveyed.

The advantages of the questionnaire are more apparent than its disadvantages, and, as a result, it frequently appeals to the amateur who uses it for all purposes regardless of its suitability and without sufficient awareness of its semi-hidden weaknesses and limitations. The major weakness of the questionnaire is undoubtedly the problem of non-returns. Not only do non-returns decrease the size of the sample on which the results are based—which is relatively unimportant wherever the sam-

ple is large—but it introduces a bias inasmuch as non-respondents can hardly be considered representative of the total population. Empirical studies have shown important differences to exist between respondent and non-respondents—and even between regular respondents and those who respond to follow-ups—in such factors as interest in the topic, attitude, conscientiousness, and promptness. An incomplete sample ordinarily includes a greater representation of the persons who are interested, who are co-operative, who are favorable to the issue under investigation, and so on. On the other hand, it is logical to assume that the non-respondents' refusal to participate is frequently not independent of such factors as a negative attitude toward the subject or toward the sponsor of the investigation. While the motives that underlie non-response vary from situation to situation, it can be assumed that the non-respondent is different, at least in some way, from the respondent. In most instances, it might be suspected that this difference may have a definite bearing on the validity of the results obtained.

The validity of questionnaire data also depends in a crucial way on the ability and the willingness of the respondent to provide the information requested. Research has shown that respondents are, as a group, of superior intellectual and educational status. Members of the lower intellectual and educational groups tend not to answer and, if they do, to introduce an element of invalidity by their inability to interpret the questions or to express their responses clearly. It also is possible that a respondent, though capable of providing the information, is not willing to divulge it. This is true especially when the information concerns sensitive subjects or reflects on the respondent, or when the respondent feels threatened by the questions asked. It also is possible for the respondent to be so uninterested in the topic under investigation that he will answer the questions more or less at random. The questionnaire frequently does not provide the investigator with sufficient opportunity for developing interest on the part of the respondent, nor does it allow him to develop the rapport necessary to permit him to ask questions of a personal or embarrassing nature. Unfortunately, the investigator has no way of knowing in how many instances both of response and of non-response the above conditions of inability and/or unwillingness to provide the information pre-

vailed, and consequently he cannot judge the extent of the invalidity of his data. His only salvation lies in selecting his population to avoid this sort of predicament.

A major disadvantage of the questionnaire is the possibility of the misinterpretation of the questions. This danger is increased when the questions are ambiguous because of improper formulation or because of the differential meaning of words associated with differences in socio-economic and cultural status—weaknesses which are as much the result of misuse as they are limitations inherent in the method itself. Misinterpretations are more likely to occur when the respondent is not equal to the task expected of him, but misinterpretations frequently arise even under ideal conditions. To make matters worse, such misinterpretations are frequently impossible to detect, nor can they be corrected as they can in the interview. Invalid responses can also occur as a result of leading questions. This weakness is not inherent in the method, however. In fact, it is less of a factor than it is in the interview, since the questionnaire is more objective. Furthermore, since the questionnaire is a matter of public record, to be scrutinized whenever unusual results occur, the presence of such a bias is more readily discernible.

Construction of the Questionnaire

Next to the choice of a suitable topic and population, probably no other aspect of a questionnaire study is more crucial to its success than is the adequacy of construction of the questionnaire itself. The average student has no concept of the complexity of devising an adequate questionnaire. His general attitude, after he has thrown a few questions together, is "Everybody knows what I mean," and it is frequently necessary to prove to him that things are not that simple. The point can sometimes be driven home by having him administer his questionnaire to a group of his colleagues and then having him analyze their responses. When, as a result of such an experience, he realizes that his questions are in need of clarification, if not complete reformulation, he is generally more receptive to suggestions for the improvement of his instrument. The difficulty is that the student confuses the questionnaire with ordinary conversation in which it is possible to correct misinterpreta-

tions through repetition of the question or further explanation. He fails to realize that once the questionnaire is in the mail, nothing can be done to improve it. He needs to appreciate that a major determinant of the quality of a questionnaire study is the adequacy of the instrument through which the data are obtained. Devising a test of intelligence, for example, might take months and even years. There is no reason to expect that a questionnaire is something that one can put together in a short afternoon.

The first step in the construction of an adequate questionnaire is to attain a thorough grasp of the field and a clear understanding of the objectives of the study and of the nature of the data needed. While a thorough review of the literature can point out the general area of significance that needs to be considered, it is usually necessary to structure the field even further, especially in an exploratory study. This is probably best done by conducting unstructured interviews with persons who are familiar with the field. Thus, in a questionnaire study of the reading interests and habits of gifted adolescents of low socio-economic status, the investigator would probably have to rely both on the literature and on interviews with such youngsters for an orientation as to what to include and how to formulate his items.

A questionnaire cannot be of infinite length. The investigator must realize that there is a limit to the demands he can make of the respondent, and that, consequently, he must limit his investigation to the point where he is not expecting too much and yet is able to get a reasonable answer to his problem. Thus, he must eliminate all questions which pertain to data which can be found readily—and often more accurately—elsewhere. If the questionnaire is still too long, he must consider what can be sacrificed with the least loss to the final answer. Every item must serve a definite purpose—or face elimination.

The more clearly the problem is stated, the more adequately each of the items can be related to the purpose of the study. This is essential not only to ensure that every item is functional, but also to encourage response, since respondents will tend to shy away from a questionnaire that is simply a fishing expedition aimed in the general direction of the target.

Most frequently, such an approach leaves unasked the very questions that would have made the study meaningful and purposeful. The following letter received by the Commissioner of Education for Alaska is, let us hope, a classic in confusion never to be equalled: however, many questionnaires display some of the same symptoms.

> I am preparing a thesis upon the subject: "The Teaching of English as Revealed in the Courses of Study of the Countries of the English-Speaking Nations of the World." I am writing you for such information and suggestions as you may be able and will kindly give me. I shall certainly appreciate whatever help you may give me along this line. Life is full of duties and we all have our own work to do; but I find that sometimes there are those who from education, training, experience and contacts in life, know off-hand what it would take a long time for others to learn from research and a long period of reading.
>
> Do you know some interesting books on Alaska: her history, her economic problems, commerce, imports, exports, human relations, religion, etc., etc.—everything of interest without our taking so much time to "think clearly" at this time.
>
> Of course, my subject is on Education and English; but these subjects require background which Alaska has.
>
> I have to present my subject in an original way, giving a new slant or fresh ideas or a definite contribution to knowledge.
>
> What is it then that Alaska has or does in a different way from other English-Speaking Countries or "outlying" parts of the United States? (May I state that we in "the States" consider you "an integral part" of the United States just as we do Hawaii. Of course you know these things. Are Alaska, American Samoa, Canal Zone, Guam, Hawaii, Philippine Islands, Puerto Rico, and Virgin Islands—all in the same class educationally as to organization? Alaska has one University and from a perusal of it, you seem to have everything.
>
> (*Marginal note*) : Could you give me the names of a few of the best books on the teaching of Englished used in Alaska? (All phases or just one branch of the work.)
>
> I wonder if climate would be the determining factor in some cases. Hawaii has her "tropical influences" in her curriculum as does the Philippines, Puerto Rico. (I always think of Cuba and want to include her in the American School System, just as I want to include Canal Zone which is a protectorate. I

am always at a loss to know what to do with Canal Zone. She only has about three cities, I believe.)

American Samoa, Guam and the Virgin Islands— (How about Wake Island?) As we think of these, what could we say of them educationally? Could we tell something interesting about them, giving their climate, area, capitals or principal cities and occupations and lead up to the need of a certain kind of education which they have or do not have and give the school census, the educational statistics including the number of teachers and the grades, classification and organization of the schools and the language existing in the islands or outlying parts and the efforts that are being made to instruct the children and the citizens or parents. What dialects have they in these "parts"?

(*Note at top of page 2:* Did your school children get a chance to see the King and Queen or hear them on the Radio?)

Of course we are interested in Alaska for your sake—and because of the "Gold Rush"—her nearness to Asia—Anthropology, —her fisheries—and I am interested in the Indians and the Esquimeaux and their carvings and also the Art of the Indians as manifested in their carvings on the totem poles, etc., etc.

I am especially interested in the railroad centers of Alaska —the cities visited by Harding and those cities made famous by the passing of Will Rogers. Keeping in mind my thesis, will you tell me something of interest about education in these cities? What Indian or Esquimeaux or other dialects have you in Alaska? I think the Americans speak the same as the Pacific Northwest or Van Couver if they are Canadians. May I hear from you? Thanks.

Very truly yours,

P.S. We are interested in Samoa because of Stevenson. I wonder if very much attention is paid to education in Samoa or Virgin Islands? I wonder what type of education is given there?[7]

Generally questions on the same sub-topic or aspect should be grouped to give the questionnaire a semblance of order, and to enable the respondent to orient himself to the trend of thought. The more general questions of a set should come first, and then the more detailed and specific—for example, "Do you work after school?" "How many hours a week do you

[7] Editor. "How Would You Answer This One?" *Alaskan School Bulletin*, 22 (1939) : 12–3.

work?" "Does your work interfere with your studies?" The questions should be arranged so that they can be cross-interpreted rather than remain completely independent, for though the questionnaire is made up of separate questions, it should be organized so that it has unity from the standpoint of purpose.

1. *Importance of Scholarly Construction.* Constructing a questionnaire calls for numerous revisions in which variations of the same question should be submitted to experimental trial. The same question posed in different ways very frequently brings out different responses. The help of outsiders is essential; they are generally more objective and can see flaws that the investigator is invariably too close to see. This points to the need for an actual pilot study, where competent persons are asked to fill out the questionnaire and to indicate their reactions to every phase of its organization. The pilot-study questionnaire can be given first to friends, then to persons who are familiar with questionnaire construction and the field in general, and finally to people of the same nature as those who are eventually to receive the final draft.

Professional people are aware of their responsibility to provide information which they feel is for a good cause. However, questionnaire studies have to compete with many other demands on the respondent's time and goodwill. Frequently resistance and annoyance toward questionnaires has been developed in potential respondents as a result of abuse by such groups as sales organizations. This only points to the challenge to be faced. If he is to expect the respondent to give of his time and energy, it behooves the investigator to prepare a questionnaire in the most scholarly fashion, for the obligation to respond to a questionnaire no longer binds when there is a legitimate question as to whether the proposed study will make a contribution. Regardless of the responsibility of the respondent, the primary responsibility lies with the investigator and the respondent's obligation to reply vanishes when the problem is trivial, when the questions display signs of carelessness, when unnecessary questions are included, when the questions do not provide the means for giving valid answers, or when the investigator has not had the courtesy to keep his demands on the respondent's time and energy within reasonable limits.

2. *Open and Closed Questions.* The form the questions and the responses are to take is an important consideration in the construction of the questionnaire. It is first necessary to determine whether the items of the questionnaire are to be *open questions,* requiring the respondent to reply in his own words—for example, "What is your occupation?"—or *closed questions,* providing the respondent with ready-made alternatives for example, in answer to the above, banker, ——, lawyer, ——, and so on. The decision is determined largely by the nature of the problem. It is generally desirable, for example, to use the open questionnaire in the early stages of investigation in order to define the field and to use a closed questionnaire when the specific aspects of the problem are more precisely delineated.

Closed questions help keep the questionnaire to a reasonable length and, thus, encourage response—and, therefore, validity from the standpoint of the representativeness of the returns—while open questions enable the respondent to give a more adequate presentation of his particular case. The open questionnaire possesses greater flexibility—which may or may not be desirable. It allows the respondent more leeway in stating his position, which may be the equivalent to saying it allows for greater validity. On the other hand, it increases the risk of misinterpretation. For example, the answer "Mechanic" in response to the question about the individual's present occupation introduces considerably greater confusion in interpretation (with corresponding loss of validity) than would listing clearly defined occupational levels for the respondent to check.

The closed questionnaire with its alternatives structures the concept under study and minimizes the risk of misinterpretation. It permits easier tabulation and interpretation by the investigator. On the negative side, the alternatives may well provide the respondent who does not have an answer with an alternative that he can check whether it applies significantly in his case or not. For example, the drop-out who leaves school for such relatively undefined reasons as "general discontent" may consider "need to work" a logically adequate, socially acceptable, and non-controversial alternative, regardless of how prominently his need to work may have featured in his decision to leave school. This is akin to the well-recognized prob-

lem of interviewer bias which is generally considered a major weakness of interview studies.

In a closed questionnaire it is essential to allow for all possible answers—that is, the categories provided must be both exhaustive and mutually exclusive. This frequently requires adding an extra category asking for "Other–please specify" for the respondent who does not find any of the alternatives provided particularly suitable. On the other hand, experience suggests that the respondent rarely exercises this option; almost invariably he simply accepts one of the alternatives provided rather than devise his own. It should be noted that the more scientifically oriented the respondent is, the more precise he tends to be, and the more annoyed he is likely to become with preplanned alternatives, each of which he would have to qualify before it would cover his particular situation.

The question of whether to use the open or the closed questionnaire can be resolved only on the basis of the usual criteria of validity, reliability, and usability, and, inasmuch as most of the problems to be covered in education are varied and complex, a combination of the two is generally better than the exclusive use of one. Each has its merits and its limitations, and it is a matter of using the proper one for the proper purpose. The closed questionnaire generally makes for greater coverage and more systematic tabulation. On the other hand, there may be the need for the respondent to clarify his position with regard to some of the items, and it is generally advisable to include an open question or two for any general reaction or comment at the end of each major section of the closed questionnaire. Neither the open nor the closed questionnaire is particularly effective for probing into a problem. When such a purpose is contemplated, the possibility of relying on the interview, particularly of the depth variety, should be considered.

The exact manner in which the respondent is to indicate his answers to a closed questionnaire depends largely on the individual questions. Certain questions can be answered by yes and no, but most answers dealing with complex aspects of a problem are not that clear-cut. The use of a five-point scale, such as *Strongly agree, Agree, Undecided, Disagree, Strongly disagree,* frequently elicits more valid responses and is less frustrating to the respondent who wants to be truthful. Whenever

the respondent is asked to rate certain items, he should be given specific directions as to the number of items he should check—for example, the *three* most important reasons—so that there will be comparability in the tabulation of the responses. If the directions simply call for "Check your favorite TV programs," the investigator would not be able to equate the responses of the person who checks only one program with those of the person who has checked a large number in which he has varying degrees of interest. In some cases, greater uniformity and, possibly, validity might be obtained by instructing the respondent to rate his favorite programs in a 1–2–3 order.

A number of rules and suggestions have been given for the construction of questionnaires.[8] These rules should be considered from the standpoint of the principles underlying scientific data-gathering rather than considered as factors peculiar to the questionnaire. The basic task is to provide a vehicle which will permit the respondent to indicate his answers truthfully and of encouraging him to do so. More specifically, the problem is one of devising an instrument of maximum validity and reliability, capable of obtaining the information relevant to a given topic. The concept of usability is also of utmost importance, since, unlike the tests administered in the captive setting of the school auspices, the questionnaire finds that its weaknesses in say, the area of excessive length, are immediately reflected in non-response and consequent loss of validity.

3. *Content.* A primary consideration in questionnaire construction is the content of the questions. Obviously, questions should be restricted to those the investigator has reason to believe will elicit valid and reliable answers. For example, the questions, "Have you noticed any improvement in the health of your child since we instituted the milk program in our school?" cannot provide usable data since the average parent cannot conceivably know. Similarly, the use of the questionnaire for the measurement of attitudes and feelings raises the very pertinent questions, "Does the person understand himself sufficiently?" and "Is he willing to reveal his interpretation of

[8] See Hadley Cantril, *Gauging Public Opinion* (Princeton: Princeton University Press, 1947), and Mildred B. Parten, *Surveys, Polls, and Samples* (New York: Harper, 1950).

himself?" Also to be avoided, for example, are questions with a "patriotic" overtone, for they are almost invariably answered in a "patriotic" direction, regardless of their content.

Each question must be justified on the basis of its contribution to the overall purpose of the study. This basic principle automatically precludes vague and ambiguous questions. Conversely, it implies clear, direct, and simple language, and general subscription to the basic rules of effective communication.

The following are samples of the type of questions to be avoided:

"What is your salary?" (9 or 12 months?); "What is the age of your father?" (He may be dead.): "Are you satisfied with your raise?" (I didn't get one.); "What is the value of your house?" (Purchase or resale?); "Do you frequently encounter disciplinary problems?" (What is frequently?); "How late do you let your children watch TV?" (Weekdays or weekends?); "Do you believe in freedom of speech?" (Emotionally toned); "Do you think a veteran should have to join a union in order to get work?" (Emotionally toned); "Do you favor federal aid to education as a means of providing for the proper education of your child?" (Emotionally toned); "Are you in favor of labor unions?" (Too broad); "Do you believe that the whole testimony of a witness found to be inaccurate in part should *ipso facto* be stricken from the record?" (Unnecessary difficulty in vocabulary); "How much income tax do you think an actress making $100,000 a year should pay?" (Lack of frame of reference); "How would you rate your superintendent?" (Lack of frame of reference); "Do you think boys and girls should be in separate classes or should they be taught together, yes or no?" (Yes or no what?); "Do you favor old-age pension and socialized medicine?" (Two ideas in one); "Marital status? ——" (Unclear; can be answered by "satisfactory"); "Boy or girl? ——" (Unclear; can be answered "yes"); "Do you make announcements at the beginning or the end of the class period? ——" (Unclear); and many other similar questions which reflect not only a lack of scholarship, but also some lack of understanding of what constitutes intelligible communication.

The plight of the poor respondent is well illustrated by Bob Burns' story of Grandpa Snazzy as a witness in court:

The attorney says: Now Mr. Snazzy, did you or did you not, on the date in question or at any time previously or subsequently,

say or even intimate to the defendant or anyone else, whether friend or mere acquaintance or in fact a total stranger, that the statement imputed to you, whether just or unjust and denied by the plaintiff, was a matter of no moment or otherwise? Answer —did you or did you not?

Grandpa thinks a while and then says, "Did I or did not what?"[9]

Questions must be worded so that they are meaningful to the person to whom they are addressed. Different expressions mean different things to different people, particularly when dissimilarities in socio-economic and cultural background are involved. The questionnaire calls for considerable educational background if it is to be answered adequately, if at all. It is generally agreed that some 40 percent of the general population are illiterate for questionnaire purposes. One authority[10] estimates that 90 percent of the people misinterpret at least 10 percent of the questions and that at least 10 percent of the people misinterpret 90 percent of the questions.

When there is reason to suspect a question is susceptible to misinterpretation, it must be phrased carefully in order to counteract possible bias. Frequently this means orienting the respondent's mind-set to the purpose of the investigation. Respondents almost invariably have already formed certain mind-sets toward a number of problems and tend to answer questions according to this frame of reference. Thus though "housework is never done," housewives usually answer "No" when asked if they work. They will frequently say "No" even when they hold a part-time job such as keeping books or minding the store for their husbands. Similarly, students will seldom list themselves as workers, though they may put in a forty-hour week in industry or business in addition to going to school. All these are things that must be foreseen and guarded against through specific questions and specific directions.

The Validation of Questionnaires

Although the criterion of validity to which the questionnaire, as an instrument of science, must subscribe has already

9 In Charles C. Ross, and Julian C. Stanley, *Measurement in Todays Schools* (Englewood Cliffs: Prentice-Hall, 1954), p. 150.

10 Mark Abrams, *Social Surveys and Social Action* (London: Heinemann, 1951), p. 74.

been defined, there remains the task of identifying the specific ways in which this validity is established. It must first be recognized that, though the whole instrument is oriented toward the whole problem, the questionnaire is comprised of specific and relatively independent questions, each dealing with a specific aspect of the overall situation. In a sense, then, it is the validity of the items rather than that of the total instrument that is under consideration. For example, the question, "How many children do you have?" may elicit a valid answer, while, in the same questionnaire, the question "How much money do you make?" can easily foster varying degrees of error, if not of deceit. It must be recognized that there are circumstances under which it is relatively impossible to obtain valid answers. Certain questions by their very nature—for example, "Do you cheat on examinations?—are likely to promote falsification. On the other hand, that the validity of the individual items must be considered does not negate the fact that the questionnaire must have a unity and validity of its own with respect to the topic under investigation. This the investigator needs to bring out through the synthesis of the responses to the specific items and an interpretation of their relevance in bringing out the total picture.

The actual validation of a questionnaire utilizes the same principles and procedures as the validation of any instrument of tests and measurements. At the most elementary level, it is necessary for the questionnaire to have face validity—that is, each question must be related to the topic under investigation, there must be an adequate coverage of the overall topic, the questions must be clear and unambiguous, and so on. A more adequate validation, however, requires checking the responses which the questionnaire elicits against an external criterion. For example, factual questions about age and educational background can be checked against the records. On the other hand, it is somewhat more difficult to locate an adequate criterion for questions of opinion and attitudes. A possible solution is to follow the questionnaire with an interview of a sample of the respondents to see whether their responses to the questionnaire actually represent their views on the subjects discussed. Similarly, a check of the validity of a grade-school child's statement that he views TV for a total of twelve hours a week might be made by having him indicate the programs which he views

regularly, by asking his parents, by having his siblings list the number of hours they view, and so on. But nothing, except a hidden monitor in the child's home, would really constitute a completely adequate criterion.

In some instances, it is possible to validate questionnaire responses against actual behavior. For example, LaPiere[11] sent a questionnaire asking hotel and motel proprietors and restaurant keepers who had actually housed and fed a Chinese couple whether they accepted Chinese guests. The responses indicated considerable discrepancy between stated policy and actual practice. This, of course, raises the question of the suitability of overt behavior as a criterion of the validity of the response to a questionnaire item. A respondent may be willing to divulge his feelings in response to a questionnaire item and yet suppress such feelings in his behavior in a face-to-face contact. Establishing validity is even more complicated in open questionnaires where the interpretation of the responses constitutes an added source of unreliability and invalidity. In some instances, the greater flexibility of the open questionnaire may promote greater validity in the responses, but it also increases the possibility of invalidity of tabulation.

Research has been conducted into the effects upon the validity of questionnaires of requiring a signature as opposed to allowing the respondent to remain anonymous. Gerberich,[12] for instance, found that requiring signatures tended to inhibit honesty and frankness in filling out the Mooney Problem Checklist. Gerberich and Mason,[13] on the other hand, found that requiring signatures made no difference in answering such questions as "Have you had a course in high-school biology?" but they warn against making an all-inclusive generalization that the identification of the respondent is irrelevant from the standpoint of the validity of his responses. This point is, of course, well taken in view of the specific nature of validity and the relatively non-ego-involved nature of the questions used in this particular investigation.

11 Richard T. LaPiere "Attitude vs. Action," *Social Forces*, 13 (December 1934) : 230–7.

12 John B. Gerberich, "A Study of the Consistency of Informant Responses to Questions in a Questionnaire," *Journal of Educational Psychology*, 38 (May 1947) : 299–306.

13 John B. Gerberich and John M. Mason, "Signed versus Unsigned Questionnaire," *Journal of Educational Research*, 42 (October 1948) : 122–6.

The validity of a questionnaire must be established prior to its use, for validation is an aspect of its development, not of its use in the solution of the problem. It should also be noted that invalidity is not restricted to the instrument itself. It can also result from systematic errors in coding or in interpretation, or from biased orientation by the cover letter or the directions.

Reliability of Questionnaire Data

The question of the reliability of the questionnaire is often ignored, partly because it is difficult to establish with any degree of precision. The usual procedures for calculating the reliability of tests are difficult to apply here. Split-half reliability is, of course, out of the question because of the relative independence and non-additivity of the component items. The possibility of phrasing the questions in two different ways and interspersing these in the questionnaire as a means of testing the reliability of certain items is of dubious validity since the average respondent would probably see through such a trick and simply ignore the second question or answer it the same way as he did the first. Besides adding to the length of the questionnaire, it is likely that such a procedure will annoy the respondent, who might think this is the type of carelessness to which he does not want to be a party, and encourage him to refuse to participate in the study—especially since it reflects on his integrity and/or his intelligence.

The *test-retest* method is the only feasible approach to the establishment of the reliability of the questionnaire. An individual who has taken the questionnaire as part of its standardization can be asked to take it again, and his answers can be compared for consistency. This procedure is not fool-proof, since on the retest the respondent will probably attempt to remember and duplicate his earlier responses rather than answer the questions as he sees them. For this reason such evidence of consistency can hardly testify to the validity of the instrument and is a questionable measure of its reliability. At the empirical level, such studies as that of Cuber and Gerberich[14] and of Gerberich[15] have shown considerable inconsistency in questionnaire

[14] John F. Cuber and John B. Gerberich, "A Note on Consistency in Questionnaire Responses," *Sociological Review,* 11 (February 1946): 13–5.
[15] Gerberich, *op. cit.*

responses, particularly in factual items, but the authors view the inconsistency to be typical of all personal communication rather than peculiar to the questionnaire. Gerberich suggests the need for the investigation of three separate but related problems: 1. the consistency of the questionnaire responses; 2. the accuracy of the questionnaire responses; 3. the comparison of the accuracy of the questionnaire responses with that of responses to the interview. He also urges the cautious acceptance of questionnaire data.

The Question of Non-Returns

Questionnaire studies are generally plagued by a relatively high percentage of non-return. Many studies in the literature report returns as low as 20 to 40 percent. Shannon[16] reports an average of 65 percent return for "reputable" questionnaire studies reported in a sample of theses, dissertations, and professional articles. He mentions, however, that a discouragingly large number of studies did not report the percentage of returns, perhaps because of inadequacies therein. On the other hand, some studies have had as high as 100 percent return. Such high returns have tended to evolve from a number of follow-ups, coupled with a happy combination of a select population and a select topic; they would be difficult to obtain in the general case.

Among the many factors that promote a high percentage of returns, none is of greater importance than the selection of a worthwhile topic and the addressing of the questionnaire to a group for whom the topic has interest and psychological meaning. No one is interested in busy work or in studies that are not likely to lead anywhere. It is the responsibility of the investigator to prove the significance of his problem to the satisfaction of the prospective respondent. Conversely, while the percentage of returns is bound to vary from topic to topic, a low percentage of returns frequently implies a poor choice of topic or of population, or perhaps inadequacy in the construction of the questionnaire—all of which can be minimized through a pilot study.

[16] John R. Shannon, "Percentages of Returns of Questionnaires in Reputable Educational Research," *Journal of Educational Research*, 42 (October 1948): 138–41.

Probably the next most important factor in promoting a high percentage of return is the follow-up. In any sample there will be a few individuals who will fail to return the questionnaire on first contact, and it is invariably necessary to institute the means for follow-up on the missing returns. In some instances, failure to return stems from a direct rejection of the questionnaire, but more frequently it implies nothing more than human forgetfulness. It is necessary, therefore, to send out follow-up letters whenever the flow of returns starts to drop off. A series of follow-ups and, finally, perhaps a double postcard calling for a brief answer to a shortened version of the questionnaire, or an interview, may be necessary to bring the returns to an acceptable level. In sending out follow-up letters, it generally is wise to include a second copy of the questionnaire in case the respondent has thrown away the first.

Of course, numerous follow-ups can be an annoyance to the respondent, leading him to refuse to co-operate in future questionnaire studies. It may also lead to his sending back results that are completely invalid. Therefore it is generally advisable in a follow-up for the investigator to attempt a new approach at convincing the potential respondent that his response is needed. The matter of follow-up is simpler when signatures on the questionnaires returned permit the identification of the delinquents to whom reminders can be sent. When signatures would be objectionable, it may be advisable to include a postcard to be mailed separately, indicating that the questionnaire has been returned under separate cover. This can be combined with the investigator's offer to mail the results of the study to those who are interested.[17]

1. *The Length of the Questionnaire.* Another significant factor in the percentage of returns is the length of the questionnaire. Generally, the shorter the questionnaire, and the less

[17] Questionnaires are sometimes coded so that the respondent, though he believes that he is given complete anonymity, can actually be identified. This might be done where, say, student reaction to a course might need to be correlated with the background of the student to be derived from his cumulative record. This procedure if fraught with danger; under no circumstances must the individual be identified *per se* except to make possible the putting together of the two segments of the information concerning him. While some people can see nothing ethically wrong with such a procedure, the student contemplating such a move should do so only after serious consideration of the matter with his advisor.

demand it makes on the respondent's time, the higher the percentage of returns. The investigator must appreciate the fact that he cannot expect his respondents to cover every aspect of a broad problem, and that he must delimit his problem to size, consistent, of course, with its retaining its meaningfulness. It must, on the other hand, be noted that the significance of the problem and the proper choice of a population, as well as the scholarship of the construction of the questionnaire, are much more important determinants of returns than is the length *per se*. Sletto[18] for example, was able to obtain a 69 percent return to a questionnaire of fifty-two pages of printed material. It would seem that brevity is not important in itself, but it is important because condensing a questionnaire frequently results in the removal of superfluous items and, thus, in a corresponding improvement in its overall quality. Although there is a rule of thumb which states that a questionnaire should not take more than half an hour of the respondent's time, the time factor must be considered in the light of the nature of the topic, the loyalty of the group contacted, and other factors—many of which are of greater importance than time itself.

2. *The Choice of Population.* The choice of the population is a prime consideration in determining the extent of response. If the topic is of interest to the respondent, he will take the time to fill out the questionnaire. Business people, for instance, certainly would respond to a questionnaire from Dun and Bradstreet. Conversely, it is likely that the low percentage of return in many questionnaire studies in education results from the fact that the questionnaire has been sent to people who do not have the answers expected of them and/or who have no interest in the subject. The population needs to be defined in such a way that participation is restricted to those who are able to make a significant contribution to the success of the study. Some investigators have been able to obtain fairly high returns by asking people in advance if they are willing to co-operate in the study, and mailing the questionnaire only to those who have indicated a willingness to participate. This is a very questionable procedure; it tends to invalidate the results before starting, inasmuch as a bias in sampling is inherent in the

[18] Raymond F. Sletto, "Pretesting of Questionnaires," *American Sociological Review*, 5 (April 1940): 193–200.

original acceptance or rejection of the request for co-operation. One is likely to get a bias soon enough without deliberately incorporating it into the design.

3. *The Instrument.* A factor not to be overlooked is the scholarship involved in the construction of the questionnaire. No one wants to be a party to slipshod work. On the contrary, if the questionnaire reflects quality, many people expect similar adequacy in the overall study and are willing to contribute to its success. Such scholarship is generally obtained at the expense of a number of revisions and, of course, pilot studies, which make possible the elimination of items that are defective, irrelevant, or otherwise objectionable. The attractiveness of the format is also conducive to higher returns; it generally pays dividends from the standpoint of returns, for instance, to have the questionnaire printed rather than mimeographed.

4. *The Cover Letter.* The cover letter or other means of contacting potential respondents is also of critical importance to the success of the study, since the investigator cannot rely on his personality to elicit co-operation, but must rely upon the printed word to "sell" his study. Sales organizations, in particular, have come to realize the crucial role of such "sales talk" in the success of any selling venture. A good letter can sell; a poor letter, on the other hand, can serve only to alienate even co-operative individuals. The letter must be brief, courteous, and forceful and also must appeal to the individual so that he will want to co-operate. The investigator might ask himself: "Specifically, what am I offering the person in return for his co-operation in the study?" Among the motives which the investigator can tap are professional obligation, personal and professional pride, spirit of helpfulness, and so on. Rarely is it adequate to base a request for co-operation on the proposition: "I have to write a thesis." Since it probably varies from study to study and from population to population, the kind of appeal that will work is probably best determined on the basis of a pilot study. It is possible, for instance, that the appeal "You will be helping to improve the situation of your fellow-students" would be effective where there was high group loyalty. In other instances, the appeal might be to the individual's personal and professional responsibility, or perhaps to the altruistic desire to help the helpless.

The cover letter should be separate from the questionnaire itself, and should be addressed to the individual by name and title. It also should bear the investigator's name and title and his relation to the study. It should make particularly clear the purpose and importance of the study, the procedure on the basis of which the respondent happened to be included in the sample, the sponsorship of the study, if any, and so on. When a student is writing to an authority in the field, the faculty advisor, as a courtesy, should also write a letter of sponsorship. Generally, adequate sponsorship promotes the study, but it may bias the responses. A study endorsed by the school board or by the steering committee of the classroom teachers association may lead teachers to go along whether they are so inclined or not. It is generally agreed that the investigator should enclose a self-addressed envelope, and it is also suggested that he include two copies of his questionnaire so that the respondent will have one for his files.

5. *Other Factors.* A number of other factors of a more minor nature frequently have a bearing on response. For instance, it has been found that the use of an ordinary stamp, rather than a prepaid stamp, promotes somewhat greater returns. Apparently people are reluctant to throw away regular stamps, but feel that a business letter stamp is not going to cost anything if it is not used. The timing appears to have an effect on returns. It is probably best not to have the questionnaire arrive on a Monday or at the beginning of the year, when the teacher or administrator is busy. On the other hand, research in this area has not been entirely consistent; it is possible that when the questionnaire deals with a topic of sufficient significance, these factors are relatively inconsequential. Perhaps it is only in instances in which the quality of the study is precarious in the first place that these factors assume significance.

6. *Dealing with Non-Response.* The matter of non-response involves two major problems. One, of course, is the maximizing of the returns in the first place. The other consideration is the adjustment of the results to compensate for non-response. For example, if the lower socio-economic subgroups responded in a much lower percentage than did the middle and upper classes, the investigator might restrict the

study to the upper and middle classes and restate his problem accordingly. Some investigators simply weight the responses of the respondents of the lower class to bring the class up to quota. This, of course, is of doubtful validity. The fact that some members of the lower classes replied while others did not suggests that the latter are really different from those who did, despite the fact that they all belong to the same socio-economic class. They cannot, therefore, be adequately represented by simple extrapolation of the respondents of their socio-economic class. Equally faulty is the scheme which takes a larger sample than is basically necessary and ignores those who do not respond. This assumes that the constant errors of sampling incorporated in such a procedure are of lesser magnitude than are the random errors—an assumption which is very questionable as we saw in Chapter 7. A more adequate scheme is advanced by Hansen and Hurwitz[19] who suggest interviewing a random sample of the non-respondents to establish their pattern of response, which can then be weighted to give an overall picture. Note that it is necessary to get the actual response of non-respondents before including them in the study through weighting. This is, of course, a much more defensible approach to the problem, but it is also much more complicated.

Evaluation of Questionnaire Research

In summary, it seems that the weaknesses of the questionnaire—while very real—are not insurmountable. It seems further that, in 1963 as in 1930, the criticisms of the questionnaire are aimed at its abuse rather than at its use. Recent opinions on the subject have run the gamut of the favorability-unfavorability continuum. On the negative side are such views as those of Charters,[20] who suggests that educational researchers must seek new ways to answer persistent questions. Even stronger positions against the questionnaire are taken by Ruckmick,[21] who knows of no other procedure which compels so much fore-

[19] Morris H. Hansen, and William N. Hurwitz, "The Problem of Non-Response of Sample Surveys," *Journal of American Statistical Association,* 41 (December 1946) : 517–29.

[20] W. W. Charters, "Beyond the Survey in School Board Research," *Educational Administration and Supervision,* 41 (December 1955) : 449–52.

[21] Christian A. Ruckmick, "The Uses and Abuses of the Questionnaire Procedure," *Journal of Applied Psychology,* 14 (February 1930) : 32–41.

thought, coupled with the avoidance of irretrievable errors as does the questionnaire; and by Duker, who writes:

> The reliability and validity is low, the frequent use of the questionnaire is a vice and a weakness mitigating against the recognition of educational research as a science. It seeks secondary information, hearsay evidence concerning facts when primary evidence is at hand. It is the voice of expediency, not of science, justified on the basis of saving time and money. It asks opinions from those not qualified to give opinions, . . . the respondent tends to put himself in the best light, and if he cannot do that he does not respond. It gives biased samples. The matter of non-response is always a question mark to the truth seeker.[22]

Frequently questionnaire research constitutes simply a pooling of ignorance, and it is conceivable that the opinion of one single expert may be far superior to the compilation of the opinions of many persons who do not know the answer. Even the Bureau of Internal Revenue, despite considerable machinery designed to enforce validity of response, has somewhat less than complete success in its use.

On the positive side are those like Monroe and Engelhart,[23] who in the 1930's suggested that until experimental science relieves us of the need of human judgment, or removes from our minds interest in unique events, this wayward child of science, the questionnaire, feeble as it is, will remain an indispensable helper. Another comment favorable to the questionnaire is that of Phillips[24] who points out that the weaknesses laid at the door of the questionnaire are primarily within the control of the investigator.

A number of studies have been reported on the relative adequacy of the questionnaire as a research instrument. Unfortunately, most of the studies have failed to point out that *adequacy* as used in this context must be spelled out according to the usual criteria of validity, reliability, and usability, and further, that validity is a specific concept. A questionnaire

22 Sam Duker, "The Questionnaire Is Questionable," *Phi Delta Kappan*, 29 (May 1948) : 386, 392.

23 Walter S. Monroe and Max D. Engelhart, *The Scientific Study of Educational Problems* (New York: Macmillan, 1936) , p. 74.

24 William M. Phillips, "Weakness of the Mail Questionnaire: A Methodological Study," *Sociological and Social Research*, 35 (March–April, 1951) : 260–7.

may be adequate for obtaining information on family size and yet not adequate for determining student reactions toward their teachers. Franzen and Lazersfeld,[25] in their study of former college students, concluded that the mailed questionnaire obtained more information and more ready admission of unusual activities and interests than did interviews.

The present consensus is that, as an instrument of science, the questionnaire has potentialities when properly used. Conrad[26] points out that the United States Office of Education makes considerable use of the questionnaire after attempting to validate it through checking returns against information on hand and various checks of internal consistency. Ruckmick[27] expresses the opinion that the questionnaire has been very useful in education and that we should not disparage it. Topp and McGrath[28] make a particularly strong plea for answering the questionnaires that one receives. They point out that the questionnaire is an economical way of accumulating information of significance to educators; that it is economical both for the sender and for the respondent in time, effort, and cost; and that if it were eliminated, progress in many areas of education would be greatly handicapped and much useful information lost. They feel that answering a questionnaire is a professional obligation, particularly since education is a profession in which there is no ready means of communication between the members, and that to say that it is not worthy of response is "to play God" and to imply that the person who sent it is lacking in common sense. They point out further that the rationalization that one does not answer questionnaires because the rate of response is generally so low that valid generalizations cannot be derived is a circular argument.

None of these statements denies the need for improvement in the questionnaire. It is fully agreed, for instance, that unless the returns can be brought up to an acceptable level, there is no point in bothering anyone. Furthermore, whenever there is

[25] Raymond Franzen and Paul F. Lazersfeld, "Mail Questionnaire as a Research Problem," *Journal of Psychology*, 20 (October 1945) : 293–320.

[26] Frank W. Banghart (ed.), *First Annual Symposium on Educational Research.* (Bloomington: Phi Delta Kappa, 1960), Ch. 2.

[27] Ruckmick, *op. cit.*

[28] Robert F. Topp and Guy D. McGrath, "About that Questionnaire—Answer it," *School Executive*, 70 (October 1950) : 59–60.

doubt as to the adequacy of the responses that can be obtained, the questionnaire should not be used. But there is, on the other hand, no justification for a complete across-the-board condemnation of questionnaire studies. The general consensus is that the questionnaire can serve a very useful and definite purpose in the advancement of education at its present stage of development—and perhaps for some time to come. It is clear, however, that there is urgent need for the improvement of its quality and for the restriction of its use to situations for which it is suited.

Evaluative Criteria

The following criteria may be used as a checklist for evaluating a questionnaire:

1. It deals with a significant topic, it makes an important contribution, and is worthy of professional participation.
2. The importance of the problem is clearly stated in the statement of the problem and in the cover letter.
3. It seeks only information not available elsewhere.
4. It is as brief as the study of the problem will allow.
5. The directions are clear, complete, and acceptable.
6. The questions are objective and relatively free from ambiguity and other invalidating features.
7. Questions that may embarrass the respondent or place him on the defensive are avoided.
8. The questions are in good psychological order.
9. The questions are so arranged that they can be tabulated and interpreted readily.

INTERVIEW STUDIES

A research method very similar in nature and purpose to the questionnaire is the interview. In fact, except for certain relative advantages which need to be clearly recognized, the two techniques are, for some purposes at least, essentially interchangeable. Although our interest is in the interview as a research technique, the interview is most frequently used in connection with non-research activities, such as counseling, the administration of an individual test of intelligence, or hiring procedures.

As a research technique, the interview is a conversation

carried out with the definite purpose of obtaining certain information by means of the spoken word. It has the same purpose and, if it is to yield dependable generalizations, must subscribe to the same criteria as other scientific techniques. It is designed to gather valid and reliable information through the responses of the interviewee to a planned sequence of questions.

The interview can be either *structured* or *unstructured*, depending on the extent to which the content and the procedures involved are prescribed and standardized in advance. Thus, in the structured interview, such as those used in the administration of the Revised Stanford-Binet, no deviation from standardization procedures is allowed. On the other hand, in a survey of the reactions of freshmen to their orientation program, a more conversational approach would allow the respondent greater freedom in discussing any aspect of the program of significance to him.

To some extent, the distinction between structured and unstructured interviews parallels that between the open and the closed questionnaire, though the unstructured interview, being even more flexible than the open questionnaire, is better suited to getting varied and sundry responses and, of course, more capable of following through on tangential ideas. Both the structured and the unstructured interview have their purpose and their relative advantages. The unstructured interview is most appropriate for getting insight into a particular situation in the early stages of investigation. The structured interview, on the other hand, is used to derive more precise generalizations in the later stages. In the structured interview, the interviewer operates on the basis of an interview schedule, which is essentially an abbreviated questionnaire, often planned to the last detail.

The structured interview calls for less versatility and on-the-spot adaptability on the part of the interviewer. On the other hand, it requires a thorough knowledge of the problem —achieved in part from the try-out of the schedule—so that the questions can be phrased to function in the field with a minimum of modification. The structured interview, therefore, can be used effectively only when a careful exploratory survey has enabled the investigator to structure the field and to devise ade-

quate questions from which deviations can be kept to an absolute minimum.

Comparison with the Questionnaire

In a sense, the interview can be considered an oral questionnaire, though, to be sure, it is more than that inasmuch as it has definite characteristics of its own which must be considered in judging its suitability for the investigation of a given phenomenon. The similarity of the interview and the questionnaire is relatively obvious in the structured interview, where the major point of distinction is that the investigation is conducted through a face-to-face contact rather than through the mails. In fact, the more structured an interview is, the more closely it resembles the questionnaire. Conversely, the less structured the interview is, the more its relative advantages and disadvantages in contrast to the questionnaire become apparent.

The primary advantage of the interview over the questionnaire is its greater flexibility which permits the investigator to pursue leads that appear fruitful, to ask for elaboration of points which the respondent has not made clear or has partially avoided, and to clarify questions which the respondent has apparently misunderstood. While the questionnaire is out of the hands of the investigator the minute it is mailed, the interview allows the investigator to remain in command of the situation throughout the investigation. The flexibility of the interview is, of course, of greatest value in exploratory studies where the field needs to be structured as the investigation proceeds. It is of correspondingly less importance where the field is more defined. For example, in the early stages of an investigation of the characteristics considered by principals and superintendents in hiring new teachers, an interview may provide a number of ideas that otherwise would be overlooked. Later, however, as more studies are made, more suggestions for drawing up a structured interview (or questionnaire) might be derived from the literature as well as from personal experience, so that the structured approach becomes more appropriate.

Despite the rigid limits it places on the interviewer, the structured interview has a number of advantages over the ques-

tionnaire under certain circumstances. It permits the establishment of greater rapport and, thus, stimulates the respondent to give more complete and valid answers; it permits the canvassing of persons who are essentially illiterate for questionnaire purposes or who are reluctant to put things in writing; and it generally promotes a higher percentage of return. Another important strength of the interview is that it permits the interviewer to help the respondent clarify his thinking on a given point so that he will give a response where he would normally plead ignorance and, even more important, so that he will give a correct answer instead of a false one. Thus if a respondent indicates that he cannot remember, the skillful interviewer may structure the field for him by pointing out some concurrent events in order to refresh his memory.

The interview also allows the observation of the respondent for signs of evasiveness, non-co-operation, and other irregularities. Not only can the interviewer appraise the sincerity and the co-operation of his respondent, but he can often combat such attitudes by establishing a higher level of rapport, or, at least, take the factor into consideration in the interpretation of the results. And, of course, the interview, by allowing for the operation of the interviewer's personality in overcoming reluctance and resistance, frequently results in successful contact with people who would refuse to participate under less compelling circumstances.

The flexibility of the unstructured interview is probably its greatest strength. Not only does it enable the investigator to pursue a given lead in order to gain insight into the problem and to obtain more adequate answers, but, more important, it frequently leads to significant insights in unexpected directions. He may, for instance, find his problem shifting as he pursues various leads and have it become an entirely different problem than the one he had anticipated. Such flexibility can also lead to by-products which were not anticipated in the original plan of the study, but which often have greater significance than the basic outcomes originally expected.

The unstructured interview is useful in probing into attitudes and motives of which even the respondent may not be aware. Depth interviewing permits getting below the level of clichés in the instance of the person who is reluctant to take a

stand, who is not too clear on his own position, or who is reluctant to admit certain things. The effectiveness of depth interviewing, of course, depends largely upon the skill of the interviewer. The psychology of projective techniques might be useful here as background for understanding the possibilities of such an approach more adequately.

Interviewer Bias

The major weakness of the interview is the interviewer bias which, ironically, stems in large part from its flexibility—which then becomes both an advantage and a disadvantage. To the extent that the interviewer is allowed to vary his approach to fit the occasion, he is likely not only to complicate the interpretation of his results, but, even more serious, to project his own personality into the situation and, thus, influence the responses he receives. Research has shown that interviewers tend to obtain data that agree with their own personal convictions. Part of this occurs as a result of ad-libbing by the interviewer in rephrasing or clarifying questions. The problem is more basic and fundamental than this, however. Psychology points out, for instance, that the very presence of the interviewer, with all that he represents in the mind of the respondent, affects the responses which he gets. This is unavoidable. Usually the respondent will orient his responses toward the sociable and the courteous rather than simply toward the truth—especially if the investigator is a pleasant person. If, on the other hand, the interviewer is curt, the respondent is likely to evade questions or even to disagree just to register his annoyance. In either case, the responses will be colored somewhat from the truth. No matter what he is or what he does, the interviewer is bound to have some effect upon his data. Research[29, 30] has shown that Negro respondents express fewer negative reactions when interviewed by white than when interviewed by Negro interviewers, and that interviewers who look Jewish or have Jewish names obtain fewer Anti-Semitic reactions than do other interviewers. While the degree of distortion present has to be appraised from

29 John Madge, *The Tools of Science* (New York: Longmans, Green, 1953), p. 244ff.

30 Herbert H. Hyman, "Problems in the Collection of Opinion Research Data," *American Journal of Sociology*, 55 (January 1950): 362–70.

the standpoint of the specific situation, this is a complication which is relatively inherent in the method itself and of which the interviewer must be fully aware, for the validity of the responses which he derives depends on his ability to overcome such biases. A considerable amount of research pertinent to the interview has been conducted in the field of clinical counseling, and the research worker contemplating doing an interview study can profit from a thorough appraisal of the literature in that field.

Another disadvantage of the interview as a research technique is its cost. Not only can it be expensive, especially when the survey covers a wide geographic area, but it is also costly in time and effort since it almost invariably necessitates call-backs, long waits, and travel. Also, a busy person may prefer to fill out a questionnaire at his leisure rather than submit to a long interview. Sometimes the advantages of the interview and the questionnaire can be combined by leaving a questionnaire to be completed and calling back at an appointed time to pick it up and to check on aspects that need clarification.

Selection of Interviewers

Contrary to the opinion commonly held by neophyte research workers, interviewing is not a technique that can be mastered on the spur of the moment. Simply talking things over with people on an off-the-cuff basis is not interviewing, and it is certainly not the scientific interviewing which is required for research purposes. On the contrary, interviewing calls for the most rigorous selection of interviewers and, further, for their most thorough, meticulous, and painstaking preparation. First, the interviewer must be a person who reflects integrity, objectivity, and personal charm, and who has the tact and ability to meet and to communicate effectively with people, even of a different cultural background. He must have a good grasp of the dynamics of human motivation and behavior and must be able to make people feel at ease and willing to communicate. He must be particularly sensitive to clues, which frequently make the difference between a successful and an unsuccessful interview, and between truth and falsehood. He must be particularly adept at making an effective primary con-

tact, for the success of the interview frequently depends on the rapport established in the first minute or two; it may even determine whether there is an interview at all. In initiating the interview, he may have to depend greatly on his friendliness and personal charm, relying on other motives, such as the interviewee's natural willingness to talk to others on subjects in which he is interested, for its continuance.

While the interviewer must be able to understand the personality dynamics of his interviewees, he must not allow them to understand him to the point of orienting their responses to what they think he would like to have them say. Furthermore, he must be aware of his own dynamics so that he appreciates that his biases sensitize him to certain phenomena and lead him to certain interpretations so that, unless he is careful, he will be looking for and seeing what he expects to see.

It is also necessary to realize that certain people just do not make good interviewers. Some cannot refrain from projecting their own personalities into the problem they are investigating, especially with respect to certain topics and certain interviewees; others do not inspire the necessary confidence in their prospective interviewees and, as a result, get an excessive percentage of refusals or are not able to keep their interviews from becoming essentially non-productive. In practice, the unsuitability of an interviewer, either in general or with respect to a specific problem or a specific type of interviewee, generally can be detected in a pilot study or in the training period conducted prior to the investigation.

Besides selecting suitable interviewers, it is necessary to fit the interviewer to the prospective interviewee. It is mandatory, for instance, that the person who interviews a housewife about some aspect of her status be sufficiently familiar with her responsibilities to permit two-way communication. Whether the interviewer should be a member of the same group, an acquaintance, or even a personal friend of the interviewee is a matter to be determined at the local level. Most experts would consider it more important for the interviewer to maintain his status as a scientific person and, except for areas of a very impersonal nature, to refrain from interviewing acquaintances where the relationship might be considered more personal

than professional and where the interviewee might be placed on the defensive. In any case, however, a pilot study is a more adequate basis for decision than *a priori* reasoning.

The interviewer must first address himself to the general task of meeting people and understanding them so that he can establish rapport quickly and effectively, overcome resistance where it develops, lead the respondent over embarrassing topics, and generally guide the conversation toward the derivation of adequate answers. In short, he must be both a psychologist and a skillful manipulator of men of varying background and status. This is even more important, of course, in the case of the unstructured interview where, unless the interviewer is particularly skillful, the conversation can go in all directions without revealing anything worthwhile, or come to a standstill. He must also know his problem and have a keen and alert mind which can detect ideas worth exploring. He must be able to help people who are inarticulate and unsure of themselves, and yet he must avoid projecting his personality into their responses.

The first task in connection with the structured interview calls for an understanding of the problem sufficient to permit the devising of adequate questions. These must be phrased in such a way as to avoid their appearing stilted when used in a conversation. Furthermore, the interviewer must be capable of deviating from the schedule to answer questions and to correct misinterpretations without violating the standardization of the instrument. This is particularly well known to people who have administered individual tests of intelligence, for instance.

Training of Interviewers

Before an interview study is undertaken, the prospective interviewers should undergo rigorous training. This is generally best done through a pilot study that will not only train the interviewers, but will also help structure the field and identify its problems and pitfalls. Training is particularly crucial in a study which involves a team of interviewers, for unless they synchronize their procedures, their findings will be essentially uninterpretable. Generally, the training program must incorporate the fourfold approach of 1. convincing the prospective interviewers that they are in need of training; 2. impressing

on them the importance of the problem to be investigated so that they will want to get valid answers; 3. orienting them to the nature of the problem so that they can see the relevance of the responses they get; and 4. providing them with special skills on the basis of which they can accomplish what they are trying to do. In structured interviews, for instance, the major task is to convince the interviewers of the necessity of abiding by standardization procedures. Supervised practice in interviewing is essential to the success of the study.

Interviewing is an art that calls for the highest level of competence—a fact which is fully recognized in such fields as counseling where the general requirements call for both a theoretical background in the dynamics of human behavior and for supervised practice in the art of interviewing. Generally, competence in interviewing comes after long years of experience coupled with a good background in the theory and the art of interviewing.[31] In view of the complex problems inherent in the use of the method, and the ease and speed with which research data can be invalidated, the use of the interview technique by the amateur—including the graduate student who has not had specific training under supervision in the field—generally is to be discouraged.

Note-Taking in Interviewing

The desirability of taking notes during an interview, in order to preclude misrepresentation resulting from a failure in memory, is relatively obvious. In some instances, this poses no problem; if the topic under investigation is such that the respondent has no objection to being quoted, his remarks can be taken verbatim or even recorded and edited later for answers significant to the study. This would be the ideal method, inasmuch as the purpose of the study may change as the analysis proceeds and certain data may assume an unanticipated significance. Many interviewees become apprehensive when they see their remarks are being recorded, however, and they become defensive, non-committal, and non-communicative. When there is danger of this occurring, it is probably best not to

[31] Survey organizations who specialize in interview studies make a serious effort to keep the same personnel over an extended period of time and to provide them with periodic in-service training.

make note-taking too conspicuous an aspect of the interview.

Whatever notes are taken should not interfere with the interview. Taking longhand notes, for instance, generally is inadvisable since it slows down the interview and is likely to encourage the respondent to become progressively more laconic. A common solution to this problem is to devise a brief interview schedule on which to check the main points of the interview according to a pre-arranged system of notation. This can be done as the interview progresses or, if any form of note-taking might be disturbing to the respondent, immediately after the interview.

Important Interview Studies

Undoubtedly the best known of the many interview studies conducted on a regular basis is the decennial census of the Federal Bureau of the Census. The regular report in which the findings of the nation-wide census are presented, and the many interim reports of the Bureau covering certain localities and aspects of the economy, are of interest to the businessman, the school administrator, and even to the average citizen. The Census is, obviously, the nation's most comprehensive interview investigation and, though it is not strictly educational, it has definite educational implications, particularly in such areas as population growth and enrollment, educational status, and income level.

Also of interest to the American public are the many polls conducted by Gallup, Roper, Crosley, and others, on a multitude of social and political issues. Polls are also conducted by a number of business firms. American Telephone and Telegraph and General Motors, for example, spend millions of dollars in questionnaires sent to their customers and patrons. General Mills and Metropolitan Life Insurance also conduct extensive polls, and radio and TV audiences are frequently canvassed by the Hooper Poll for the purpose of deriving Hooper ratings as an index of the relative popularity of the various programs.

The Kinsey studies,[32, 33] of the sexual habits of American

[32] Alfred C. Kinsey, *et al., Sexual Behavior of the Human Male* (Philadelphia: Saunders, 1948).

[33] Alfred C. Kinsey, *et al., Sexual Behavior of the Human Female* (Philadelphia: Saunders, 1953).

males and females are of interest here because they exemplify some of the difficulties involved in conducting sociological research into areas which are of a confidential and personal nature. The major criticism of the Kinsey studies—in addition to loose reporting—centers around the problems of sampling and of interviewing, both of which, because of the nature of the problem involved, introduce special difficulties. These studies have been reviewed by numerous critics, and the student is referred to more comprehensive sources for more adequate treatment of their net worth.[34, 35]

Validity of Interview Studies

Establishing the validity of the interview presents much the same problems as it does for the questionnaire. Again, validity pertains to the separate items as well as to the overall technique. The fact that the interview permits following through on misunderstood items and inadequate responses generally promotes validity, but suitable criteria, especially for the more sensitive and intangible issues, are relatively unavailable. The rather common practice of using inter-interviewer agreement as a criterion of validity is questionable in view of the inherent danger that, if interviewers have the same frame of reference because of similarity in background and training, they may simply duplicate each other's mistakes.

A crucial point in the validity of the interview is the possibility—if not the likelihood that the interviewer's very presence will affect the responses which he gets. Unless special care to avoid such a bias is exercised, the results can be misleading. The validity of the interview appears to be directly proportional to the competence of the interviewer. This makes its use by the amateur in any but the most psychologically simple situations relatively precarious, and the method, though most valuable when properly used, should be approached cautiously.

The reliability of the interview also must be considered from the standpoint of the individual items, and, while it may

34 W. Allen Wallis, "Statistics of the Kinsey Report," *Journal of American Statistical Association*, 44 (December 1949) : 463–84.

35 William G. Cochran, *et al.*, "Statistical Problems of the Kinsey Report," *Journal of American Statistical Association*, 48 (December 1953) : 673–716.

be possible to obtain reasonable consistency in certain items, a similar consistency can hardly be expected in other matters. This is probably not peculiar to the interview, however, but would be true of any approach used to obtain the same data.

SUMMARY

1. Survey studies are oriented toward determining the present status of a given phenomenon, and, though too frequently they are restricted to semi-clerical fact-finding expeditions conducted under relatively ill-defined circumstances, they are often of considerable immediate value. They can also provide as by-products, an indication of trends and even hypotheses as to the antecedents of the status noted. Their flexibility makes them particularly suited to the early exploration of phenomena. They are, however, of relatively limited scientific sophistication. There is need for a greater utilization of survey results as sources of hypotheses and for greater emphasis on the interpretation and integration of the findings into theoretical structure.

2. Although the distinction is not clear-cut, surveys can be divided into descriptive and analytical. The questions of sampling and of the validity of the various data-gathering instruments are of crucial importance to the validity of all survey results.

3. Survey testing probably represents the most systematic research program conducted in our schools. The proper interpretation of the results of survey testing requires considerable background in the field of tests and measurements, especially from the standpoint of the validity of the instruments used in the particular situation.

4. The questionnaire is probably the most used and the most abused survey instrument. Too frequently, it is used to provide a pooling of ignorance in situations where only a more adequate approach—experimentation, for example—can provide a meaningful answer. The question of non-returns is particularly troublesome since non-response generally introduces a bias in the data. The possibility of misinterpreting the items is another source of difficulty relatively inherent in the questionnaire method. On the other hand, it has obvious advantages, particularly from the standpoint of practicality.

5. Among the more important considerations in the successful use of the questionnaire in educational research are the appropriateness of the questionnaire to the investigation of the particular problem; the worthwhileness of the problem and the proper choice of the population; and the scholarliness of the instrument and the appeal of the cover letter.

6. The questionnaire can be open or closed, or it can combine the two approaches, depending on the nature of the problem and

the purpose of the study. The open questionnaire, for example, is more flexible and is generally better suited for the early exploration of a problem.

7. Evaluations of the questionnaire range from outright condemnation to general endorsement. It is generally agreed that there is a need for its overall improvement and the restriction of its use to situations where it is appropriate.

8. Although for certain purposes, the interview is interchangeable with the questionnaire, it has definite characteristics—and advantages and disadvantages—of its own. The unstructured interview, for example, is particularly flexible and, therefore suitable for the early stages of a problem. Its weakness lies in the bias which the very presence of the interviewer is likely to introduce in the data which he collects. The rigorous selection and training of the interviewers is essential to the success of the interview.

9. Note-taking during the interview is sometimes a problem; the danger of distortion and omissions resulting from memory losses must be balanced against the distortion which may result when the interviewee realizes that his responses are being recorded.

PROJECTS and QUESTIONS

1. Evaluate the results of the academic testing program of a given school system. What conclusions do the data warrant? Specifically what steps were taken to improve present status?
2. What is wrong with the survey approach to the investigation of educational problems?
3. List a few problems for which the questionnaire would be a legitimate tool of investigation.
4. Locate a questionnaire in the literature and appraise its quality in the light of the principles of test construction.
5. As a class project, prepare and pretest a questionnaire. Include a cover and a follow-up letter.

SELECTED REFERENCES

ABRAMS, MARK. *Social Surveys and Social Action*. London: Heinemann, 1951.

ACKOFF, RUSSELL L. and PRITZKER, LEON. "The Methodology of Survey Research," *International Journal of Opinion and Attitude Research*, 5 (Fall 1951): 313–34.

HYMAN, HERBERT H. *Survey Design and Analysis: Principles, Cases, and Procedures*. Glencoe: Free Press, 1955.

The Questionnaire

ALBRECHT, GUSTAV. "A Survey of Teacher Opinion in California," *Phi Delta Kappan*, 42 (December 1960): 103–8.

BLANKENSHIP, ALBERT B. "Pretesting a Questionnaire on Public Opinion," *Sociometry*, 3 (July 1940): 263–9.

BROWN, WILLIAM F. and HOLTZMAN, WAYNE H. "A Study-Attitude Questionnaire for Predicting Academic Success," *Journal of Educational Psychology,* 46 (February 1955): 75–84.

CAMPBELL, ALBERT A. "Two Problems in the Use of the Open Question," *Journal of Abnormal and Social Psychology,* 40 (July 1945): 340–3.

CLAUSAN, JOHN A. "Controlling Bias in Mail Questionnaires," *Journal of the American Statistical Association,* 42 (December 1947): 497–511.

DI VESTA, FRANCIS J. "Problems in the Use of Questionnaires for Studying the Effectiveness of Educational Programs," *Educational and Psychological Measurement,* 14 (Spring 1954): 138–50.

EL-BADRY, M. A. "A Sampling Procedure for Mailed Questionnaires," *Journal of the American Statistical Association,* 51 (June 1956): 209–27.

FISCHER, ROBERT P. "Signed versus Unsigned Questionnaires," *Journal of Applied Psychology,* 30 (June 1946): 220–5.

FRANZEN, RAYMOND and LAZARSFELD, PAUL F. "Mail Questionnaires as a Research Problem," *Journal of Psychology,* 20 (October 1945): 293–320.

HEBERLE, RUDOLPH. "On the Use of the Questionnaire in Research: Open Letter to a Graduate Student," *American Sociological Review,* 16 (August 1951): 549.

KOOS, LEONARD V. *The Questionnaire in Education.* New York: Macmillan, 1928.

LENTZ, THEODORE F. "Reliability of Opinionnaire Technique Studied Intensively by the Retest Method," *Journal of Social Psychology,* 5 (August 1934): 338–64.

MORREN, ROBERT L. and ROTHNEY, JOHN W. M. "Personalizing the Follow-Up Study," *Personnel and Guidance Journal,* 34 (March 1956): 409–12.

NATIONAL EDUCATION ASSOCIATION (Research Division). "The Questionnaire," *N.E.A. Research Bulletin,* 8 (January 1930): 1–51.

NIXON, JOHN E. "The Mechanics of Questionnaire Construction," *Journal of Educational Research,* 47 (March 1954): 481–7.

NORMAN, RALPH D. "A Review of Some Problems Related to the Mail Questionnaire Technique," *Educational and Psychological Measurement,* 8 (Summer 1948): 235–47.

PARTEN, MILDRED B. *Surveys, Polls, and Samples.* New York: Harper, 1950.

PAYNE, STANLEY L. "Thoughts about Meaningless Questions," *Public Opinion Quarterly,* 14 (Winter 1950–51): 687–96.

PHILLIPS, MARJORIE. "Problems of Questionnaire Investigation," *American Association of Health, Physical Education, and Recreation Research Quarterly,* 12 (October 1941): 528–37.

PHILLIPS, WILLIAM M. "Weakness of the Mail Questionnaire: A Methodological Study," *Sociological and Social Research,* 35 (March–April 1951): 260–7.

REID, SEERLEY. "Respondents and Non-Respondents to Mail Questionnaires," *Educational Research Bulletin,* 21 (April 1942): 87–96.

ROMINE, STEPHEN A. "Criteria for a Better Questionnaire," *Journal of Educational Research,* 42 (September 1948): 69–71.

ROSEN, HJALMAR and ROSEN, R. A. HUDSON. "The Validity of 'Undecided' Answers in Questionnaire Responses," *Journal of Applied Psychology,* 39 (June 1955): 178–81.

SCATES, DOUGLAS E. and SCATES, ALICE Y. "Developing a Depth Questionnaire to Explore Motivation and Likelihood of Action," *Educational and Psychological Measurement,* 12 (Winter 1952): 620–31.

SLETTO, RAYMOND F. "Pretesting of Questionnaires," *American Sociological Review,* 5 (April 1940): 193–200.

SLOCUM W. L., *et al.* "Increasing Response to Questionnaire and Structured Interviews," *American Sociological Review,* 21 (1956): 221–5.

SUCHMAN, EDWARD A. and MCCANDLESS, BOYD. "Who Answers Questionnaires?" *Journal of Applied Psychology,* 24 (December 1940): 758–69.

TOOPS, HERBERT A. "Questionnaires," in Monroe, Walter S. (ed.), *Encyclopedia of Educational Research.* New York: Macmillan, 1950.

VOAS, ROBERT B. "A Procedure for Reducing the Effects of Slanting Questionnaire Responses toward Social Acceptability," *Educational and Psychological Measurement,* 18 (Summer 1958): 337–45.

WALLACE, DAVID. "A Case For and Against Mail Questionnaires," *Public Opinion Quarterly,* 18 (Spring 1954): 40–52.

The Interview

BINGHAM, WALTER V. and MOORE, BRUCE V. *How to Interview.* 4th ed.; New York: Harper, 1959.

BLANKENSHIP, ALBERT B., *et al.* "Survey on Problems of Interviewer Cheating," *International Journal of Opinion and Attitude Research,* 1 (March 1947): 63–77.

CALLAHAN, DON, *et al.* "Interviewer Bias Involved in Certain Types of Opinion Survey Questions," *International Journal of Opinion and Attitude Research,* 1 (March 1947): 63–77.

COCHRAN, WILLIAM G., *et al.* "Statistical Problems of the Kinsey Report," *Journal of the American Statistical Association,* 48 (December 1953): 673–716.

CRESPI, LEO P. "The Interview Effect in Polling," *Public Opinion Quarterly,* 12 (Spring 1948): 99–111.

DEANE, STEPHEN R. "The Interview as a Tool of Adult Education Research," *Adult Education Bulletin,* 14 (June 1950): 150–7.

HYMAN, HERBERT H. "Problems in the Collection of Opinion-Research Data," *American Journal of Sociology,* 55 (January 1950): 362–70.

KANN, ROBERT L. and CANNELL, CHARLES F. *The Dynamics of Interviewing: Theory, Techniques, and Cases.* New York: Wiley, 1957.

KOGAN, LEONARD S. "The Electrical Recording of Social Casework Interviews," *Social Casework,* 31 (1950): 371–8.

MAULDIN, W. PARKER and MARK, ELI S. "Problems of Responses in Enumerative Surveys," *American Sociological Review,* 15 (October 1950): 649–57.

MOSTELLER, FREDERICK. "Interviewing and Questionnaire Design," in *The Pre-Election Polls of 1948.* New York: Social Science Research Council, 1949.

PARKER, CLYDE A., *et al.* "Questions Concerning the Interview as a Research Technique," *Journal of Educational Research,* 51 (November 1957): 215–21.

WALLIN, PAUL. "An Appraisal of Some Methodological Aspects of the Kinsey Report," *American Sociological Review,* 14 (April 1949): 197–210.

WALLIS, W. ALLEN. "Statistics of the Kinsey Report," *Journal of the American Statistical Association,* 44 (December 1949): 463–84.

YOUNG, PAULINE V. *Scientific Social Surveys and Research.* Englewood Cliffs: Prentice-Hall, 1949.

. . . the history of science is a history of relentless analysis. We aim to break down gross phenomena into subphenomena.

BENTON J. UNDERWOOD

11 The Survey: Analytical Studies

Analysis as an Aspect of Science

One of the most fundamental of all research techniques is *analysis.* Fundamentally, analysis is a method which underlies the whole process of research, from the selection of a problem and its reduction in size to the point where the data are processed and the conclusions are reached. Since most educational problems are too broad to be attacked as a unit, they must be analyzed into their constituent parts as the preliminary step to deriving significant relationships among them, to isolating relevant from irrelevant aspects, and to structuring them in their scientific contexts.

As we have seen, an aspect of the early development of science is *classification*—a process which consists of the analysis of phenomena into their basic components in order to isolate whatever properties will prove relevant as the basis for ordering them into sub-categories meaningful for the purpose at hand. The purpose of classification as a technique of science is to allocate phenomena into homogeneous sub-classes for which more

precise relationships can be discovered. In this process, analysis plays the critical role of identifying the crucial aspects of phenomena, and thus not only provides a greater understanding of both the whole and the parts, but also permits the allocation of phenomena into ever-more precise and functional categories. It also permits combining into meaningful classes phenomena having in common one or more aspects crucial with respect to a given purpose, despite perhaps vast differences in irrelevant aspects. Thus, to a zoologist, a whale is a mammal, not a fish.

Such breakdowns are essential to the development of a discipline into a precise science. Generally the properties of a given phenomenon are predicated on the properties of its constituents, and the identification and understanding of these constituents, at one level or another of analysis, can usually lead to an understanding—or at least to hypotheses—of the nature of the phenomenon itself. Thus, the identification of a child as an under-achiever provides considerable insight into many of his characteristics and his likely behavior under certain circumstances.

The point in the analysis at which the breakdown provides the greatest enlightenment depends on the purpose. Thus, the analysis of materials into the basic elements of the periodic table constitutes one of the greatest "discoveries" in the advancement of science; the more recent and further analysis of matter into its atomic structure constitutes an even more fundamental step in its progress. On the other hand, neither breakdown is too useful in understanding the nature and properties of table salt ($NaCl$), which does not have much in common with the properties of its constituent elements nor of its constituent atoms.

Of course, analysis can be carried too far for the purpose under consideration, and the investigator must decide on the degree of fineness with which he wants to analyze his data. The point to consider here is that, for certain purposes at least, a given phenomenon loses its meaningfulness if it is dissected past the point at which it really exists as an entity, and that the researcher must stop short of complete analysis or face the risk of destroying the very thing he is investigating. This is essentially the objection that Gestalt psychologists raise against

analysis in their basic statement that the whole is more than the sum of the parts. The analysis of a phenomenon into its constituents provides a greater understanding of its nature only to a point. Beyond that, the basic laws which apply to the phenomenon itself may no longer apply to its constituents.

Analysis is worthwhile only to the extent that the breakdown is relevant from the standpoint of the study. Thus, while the analysis of a document with respect to such components as "appeal to emotion," "appeal to logic," and so on might provide increased insight into the psychology of its appeal, its analysis on the basis of vocabulary might be more appropriate for the purpose of appraising its readability and predicting likely success or failure in its comprehension. The analysis of the document into the letters that form the words would probably provide little of any value in the usual case.

Analysis as a Research Method

In addition to being a fundamental method of science, analysis is also a legitimate research method in its own right. It is particularly closely related to descriptive research with which it plays an essentially complementary role. Not only is analysis really a form of description, but without analysis to provide a deeper insight into their basic nature, the adequate description of phenomena is relatively impossible. The description of the nature of a textbook, for example, can be only superficial without some attempt to analyze its various characteristics.

Analytical research—frequently called *content analysis* or *documentary analysis*—is generally associated with the analysis of the content of speeches, textbooks, editorials, TV programs, or, perhaps, essay examinations from the standpoint of prejudice, readability, nature of the mental processes involved, and so on. At a more sophisticated level, content analysis may involve textbook analysis, job analysis, and factor analysis. Job analysis attempts to analyze the nature of a job in order to permit a more adequate allocation of the worker to the job. Job or activity analysis in the field of education might comprise time-and-motion studies of the duties and responsibilities of school personnel, from the superintendent to the janitor. Such studies would be particularly valuable to the administra-

tor in selecting personnel and in providing in-service training for meeting job requirements. They would also be helpful to teacher-education institutions in providing students with the skills required to fill the position for which they are being trained.[1] In all instances of such analytical research, the purpose is to identify significant factors along the dimensions into which phenomena can be categorized from the standpoint of the problem at hand. In analytical studies particular emphasis is placed on the identification of the relationships among the various aspects of the phenomena, inasmuch as relationships among components are frequently more important than the components themselves. Content analysis is of considerable value to education both in the derivation and revision of the curriculum and in the understanding of some of the complex variables encountered in the field.

DOCUMENTARY-FREQUENCY STUDIES

A form of content analysis of particular interest to educators is that used in *documentary-frequency* studies, which are used to determine the frequency of occurrence of certain phenomena. For example, an investigator might undertake to determine the common vocabulary of children through an analysis of their letters, themes, and other writings. Similar studies have been done of such topics as the nature of fractions in common use, the errors that are committed in the various academic skills, and so on. Among the classic studies in this category, probably none is more well known than the pioneer vocabulary studies conducted by Thorndike in which he identified the 10,000, 20,000, and 30,000 most used words.[2]

Particularly interesting are the studies of readability, which are aimed at determining the level of reading difficulty of written material. The study is usually accomplished by means of a formula, a number of which have been devised, each based

[1] In this category, one could also include *case analysis* whose purpose is to identify conditions in the individual's past and present circumstances that might have been involved in a causative or contributing way in the development of his present predicament or status. Case studies will be considered in Chapter 12 since they probably have greater bearing on research involving the discovery of the antecedents of phenomena than they do on an analysis of their present status.

[2] See Chapter 15.

on somewhat similar and yet different bases, and yielding essentially similar but not identical results.[3] Such studies have shown, for example, that textbooks frequently have a measured reading level beyond the grade placement for which they are prescribed.[4]

Textbooks also can be analyzed from the standpoint of any number of aspects, such as emphasis on group discussion, inclusion of complex terms, apparent advocacy of socialistic ideologies, treatment of minority problems, use of graphic material, format, and so on. An important study of this nature is Smith's *One Hundred and Fifty Years of Arithmetic Textbooks*[5] which analyzes arithmetic textbooks written since the early 1800's from the standpoint of content, methods, and problems.

A more comprehensive approach to textbook analysis might involve their investigation on a multiple basis. For example, in selecting social-studies textbooks for adoption by the school, the investigator might base his analysis on such factors as historical accuracy, exposition of acceptable social ideals, emphasis on character formation, readability, attractiveness of format, motivational appeal, clarity of expression, suggestions for use and availability of teaching aids, continuity of the series, and so on. The analysis could be extended to a rating and a weighting of each of the factors in proportion to its alleged importance from the standpoint of the objectives of the school curriculum. This would then yield an overall score or index.

Documentary-frequency studies can be particularly valuable in curriculum revision. While they are concerned primarily with present status, they are definitely oriented toward the improvement of future practice. On the other hand, frequency studies contribute rather little to the development of education as a science. Too frequently, factors of availability

[3] See Irving D. Lorge, "Readability Formulae: An Evaluation," *Elementary English*, 26 (February 1949) : 86–95; Edgar Dale and Jeanne S. Chall, "A Formula for Predicting Readability," *Educational Research Bulletin*, 27 (January, February 1948) : 11–20, 37–54.

[4] George G. Mallinson, *et al*, "The Reading Difficulty of Some Recent Textbooks for Science," *School Science and Mathematics*, 57 (May 1957) : 364–6.

[5] Henry L. Smith, *et al*, *One Hundred and Fifty Years of Arithmetic Textbooks*, School of Education Bulletin, 21: No. 1. (Bloomington: Indiana University, 1945) .

and ease of measurement have oriented such studies toward the investigation of the trivial and resulted in the neglect of the more fundamental aspects of phenomena. For example, the factors of experience and training are generally considered in the study of teacher effectiveness, while personality factors, which may have an even greater bearing on effectiveness, are frequently ignored because they are more difficult to appraise. It must also be realized that such studies do not identify the reasons why the phenomena discovered actually exist; they simply point to their existence.

Caution must be exercised in the interpretation of the results of documentary-frequency studies. While they can provide valuable information to be considered in the revision of the curriculum, for instance, it must not be assumed that frequency of occurrence of a given phenomenon is synonymous with its importance. The fact that adults make little use of fractions of the variety of 7/19 implies that they should be eliminated from the curriculum only if we assume that what is not used is not useful nor necessary. Similarly, a study revealing that a certain amount of duplication exists in the curriculum does not answer the question of how much duplication is permissible, or even desirable. This is, of course, not peculiar to documentary-frequency studies; research is never expected to provide decisions, but simply data on which intelligent decisions can be based.

Another interesting aspect of documentary frequency studies is the tediousness involved. Thorndike's 10,000 most used words, for example, were selected from over 7,000,000 words found in forty-one different sources.[6] Fortunately, with the advent of the modern electronic computer, this task is becoming relatively clerical. Not only can we sort the material on the punched-card system, but we can now also read the information on the magnetic tape of the computer and have the data classified alphabetically and frequencies tabulated.

OBSERVATIONAL STUDIES

Nature of Observation

Observation is at once the most primitive and the most refined of modern research techniques. It is, undoubtedly, the

[6] Thorndike, *op. cit.*

first procedure of science, inasmuch as all scientific data must originate in some experience or perception. As a scientific tool, it may range from the most casual and uncontrolled to the most scientific and precise, involving modern mechanical and electronic means of supplementing observation. Much of the observation of the layman is of a capricious nature, and it generally does not yield results of any great scientific significance. It differs from scientific observation only in degree, however, and there is no point at which observation ceases to be nonscientific and becomes scientific. On the contrary, observation can be made progressively more scientific to meet the needs of the particular situation, and observation is a fundamental tool even at the most advanced levels of science.

Observation underlies all research; it plays a particularly prominent part in the survey procedures now being considered, but even experimentation is simply observation under controlled conditions. As a research technique, however, observation has made but relatively limited contribution to the development of education as a science. Thus far, most of the uses of observation as a research technique—with the obvious exception of child development studies—have been relatively routine and scientifically imprecise. On the other hand, it must be recognized that many significant variables can be investigated in no other way.

Criteria of Scientific Observation

Contrary to the opinions of the amateur, who may think of observation as something that anyone can do, observation is so loose and yet so complex that it is frequently one of the most difficult techniques to harness in the service of science. It is, therefore, necessary to make a distinction between observation as a scientific tool—which must *ipso facto* comply with the usual requirements of all instruments of science—and the casual observation of the man in the street. Both the scientist and the layman observe, but the scientist starts with a hypothesis and arranges the conditions of his observations to avoid distortion. More specifically, scientific observation must comply with the following criteria:

1. Scientific observation is systematic rather than haphazard or opportunistic. Although in the early stages, where the prob-

lem is to survey the phenomenon as a whole, it is necessary to maintain maximum flexibility in order to gain insight into its nature and to permit structuring the field for more controlled investigation later, in the more refined stages at which research operates, casual observation rarely provides anything of value. Scientific observation is directed at those specific aspects of the total situation which are assumed to be significant from the standpoint of the purpose of the study. The layman, on the contrary, frequently overlooks what is crucial while he devotes his attention to what is irrelevant.

The scientific observer is an expert who knows precisely what he is looking for in the total situation. On the basis of his familiarity with the phenomenon, he stages the observation to isolate those aspects of the situation which are significant for his hypothesis. He not only structures the phenomenon he is to observe, but he also plans his observations to prevent his overlooking significant aspects. He is aware of the pitfalls to be avoided and he has the background of experience, both in research and in the problem area, necessary for him to capitalize on the opportunities that present themselves.

Scientific observation is based on the assumption that orienting the observation toward narrow bands leads to greater dependability of observation. This, in turn, assumes that the phenomenon has been properly dissected and its significant aspects correctly identified. There is, of course, the inherent danger of overlooking significant components of the situation, since any attempt at orienting observation produces a mind-set that is likely to blind the observer to other aspects of the situation. It is also essential that the categories used as a basis for orienting observation be neither so broad that they lead the investigator to see nothing but relative confusion nor so narrow that they rob the components of their significance.

2. Scientific observation must be as objective and free from bias as possible. This must be reconciled with the fact that scientific observation—like all research—should generally be guided by a hypothesis. Again, we can raise Bacon's objection to the dangers of hypotheses in directing the investigator's search toward preconceived goals. Undoubtedly, prejudgment on the part of the observer may color his perceptions and blind him to certain aspects of the actual situation. The teacher who is convinced that under-achievers are "lazy" is

likely to find many confirming instances. Prejudgment is a liability, particularly in the early stages of observation where the observer must maintain maximum flexibility and open-mindedness as to what is relevant and crucial from the standpoint of his purpose.

On the other hand, it is unrealistic to expect an investigator to begin even preliminary operations without some idea as to what he is likely to find. Furthermore, it does not seem desirable that he do so. Although it is possible that a hypothesis may orient the investigator in the wrong direction, he cannot deal adequately with a complex situation if he simply looks at everything on an opportunistic basis. The investigator should realize that his perceptions will be influenced by his experiences and openly acknowledge his basic premises as working hypotheses—all the while relying on the scientific method and the restrictions it imposes on the operation of his judgment to minimize the influence of such predispositions. This is, of course, a more difficult task because of the highly subjective nature of observation.

The observer must, of course, maintain his neutrality: not only must he consider hypotheses as something to be tested rather than proved, but he must, at all times, maintain a flexible attitude so that he can deviate from his original plans when such deviation appears advisable. This frequently means that he will have to fight his whole background; it may even mean that under certain circumstances and conditions, he may have to disqualify himself, just as judges occasionally do in cases in which they feel they cannot be impartial. A person with a high level of repressed hostility, for example, might not be able to conduct an impartial study of the disciplinary measures used in the classroom.

3. The observer must be in a good position to observe and he must have adequate sense organs. He must have a clear conception of the overall, as well as the specific, aspects of the situation and be able to distinguish the significant from the insignificant. He must be alert to what he sees and be able to make adjustments on the spur of the moment. And while he must be systematic and objective in his observations, he must also display originality, flexibility, and imagination.
4. Wherever possible scientific observation should be quantitative. Although many important phenomena cannot be quantified, it becomes almost imperative in the more refined stages of investigation to derive some means of quantifying

observations in order to increase their precisions and to facilitate their analysis.

5. Like all scientific data-gathering techniques, observation must comply with the usual criteria of reliability, usability, and, especially, validity. While these characteristics have to be appraised from the standpoint of the individual case, they suggest, among other things, the need for a number of observations covering a relatively large segment of the phenomenon under study. It is recognized in anecdotal records, for example, that the teacher must be careful not to report the atypical behavior of the child and, thus, present a misleading picture of his true nature. Furthermore, since scientific observation must be verifiable, it must be carefully recorded so that verification becomes possible.

Observation as a Scientific Procedure

Unlike the questionnaire and interview which rely on the respondent to provide the data required, observation allows phenomena to reveal themselves through their operation or characteristics. Observation is the most direct method of collecting certain data, since it attempts to derive the data directly rather than through the reports of the individuals involved. On the other hand, such phenomena as attitudes cannot be observed; their existence must be inferred from behavior. Observation is particularly useful in situations involving infants who are unable to verbalize. Observation is also valuable in situations which are inherently complex, such as a study of democracy in action as represented by the resolution of a social issue through discussion. The use of observation as a research technique would also be indicated in investigations of animals and inanimate objects and many other phenomena, which can be investigated only through observation, no matter how imprecise such investigations may be.

Among the advantages of observation is that it permits the recording of behavior as it occurs and eliminates having to rely on the reports of untrained observers. The scientist is generally more accurate in his observations and is more likely to know what to look for than the layman. Observation has the further advantage over such techniques as the questionnaire and the interview that it may not require the same degree of co-operation by the subject. The observer's obvious limitation in this respect is the relative unlikelihood of his being on the spot when sig-

nificant phenomena occur so that he is more or less forced, for the sake of economy, to rely on questioning people who happen to have been there. It is sometimes possible for him to set the stage for the occurrence of certain phenomena, but in such cases, he must be careful not to introduce an element of artificiality into the situation and thus invalidate his observations.

The Training of Observers

Observation is no better than the people who use it. It is, therefore, mandatory for anyone who conducts an observational study to undergo extensive training in order to ensure validity and reliability in his observations. Sportscasters and announcers, for example, are much more adequate observers of a sports event than is the average fan; not only are their observations more accurate and dependable, but they also see many things that escape the layman. They can anticipate plays and, knowing what is likely to occur, they can orient their total observation to the significant components, while the fan simply scatters his attention on irrelevant aspects. Training in observation is particularly important in the social sciences where, because many of the situations to be observed are highly complex, the observer is frequently faced with determining which factors are significant out of the multiple phenomena occurring simultaneously.

It must also be remembered that some people are just not good observers, either in general or with respect to certain phenomena in which they may be emotionally involved or otherwise unsuited. Observers should be carefully selected and, in the course of their training, a check should be made of their suitability for the observation of the particular phenomenon under investigation.

Planning for Observation

Securing valid observations demands careful planning, for unless the observer is oriented to the purpose and the crucial aspects of a situation, his observations will probably have no more validity than those of the blind men looking at the elephant. Accuracy of observation requires first that the observer be in a favorable position for observing. It is also necessary to decide the extent of the individual observations. It may be,

for instance, that the observation of a tangible object requires only a quick look. The adequate observation of complex intangibles calling for inferences about their nature—for example, evidence of repressed hostility among teachers in the classroom—may, on the other hand, require extended observation and even the pooled judgments of many observers. Such problems are generally best resolved on the basis of a pilot study that permits the observer to appraise the situation beforehand to determine how he can best observe without distortion or oversight of significant aspects.

An important consideration in planning for observation is the adequacy of the sample of observations on which the conclusions are to be based. This calls for the dual process of deriving a representative sample of observations from a representative sample of the subjects under investigation—with "representativeness" in both cases defined from the standpoint of the purpose of the study. For example, in the determination of children's behavior patterns, a representative sample of children needs to be placed in an observational situation in which they are likely to display their typical behavior. If, on the other hand, the problem is to determine the reactions of delinquent children to conditions of stress, one would have to choose, not a random sample of children, but a random sample of delinquent children, and further to expose them not to ordinary conditions but rather to preplanned conditions of stress. It would be incorrect, for example, to base conclusions concerning the behavior of school children on observations obtained at 2:30 in the afternoon when children are more likely to be listless and tired. (Nor would it be acceptable to rely on volunteers, as we saw in Chapter 7.)

The observer needs to know beforehand the type of observations he is to make—whether he is simply to note the occurrence of certain events on a yes-no basis, or whether he is to make a judgment as to their intensity, duration, and apparent effect. It also must be clearly understood by the observer whether he is simply to observe or whether he is to interpret what he observes. For example, is he to note simply that John shoved his neighbor, or is he to relate the behavior to such psychological dimensions as hostility, social immaturity, and so on. If he is to make interpretations, it is essential that he

know the criteria on which he is to base his judgments. Furthermore, regardless of what is involved, plans must be made for recording the information quickly, in order to keep distraction of the observer to a minimum, and inconspicuously, in order to prevent distortion of the situation.

Science has developed a number of instruments and devices of various degrees of scientific sophistication designed to promote more precise observation. While none of the instruments used in the social sciences has achieved the precision and accuracy of the gauges, meters, and other yardsticks of the physical sciences, we do have motion and still pictures, sound-recording equipment, one-way screens, projectors, and various psychological scales, as well as the simple checklist. Because they can be stopped at any moment or played any number of times and even in slow motion without distortion, movies are particularly useful in observing a complex situation, such as the operation of democratic discussion in a large group. Instrumentation can also be valuable when used in connection with observation in a laboratory situation where, because of the restrictions placed on what is being observed and the control of irrelevant factors, such observation can be made relatively precise. Scientific progress has been made in the area of reading, for instance, where instruments have made possible the derivation of relatively dependable generalizations concerning eye movements and the other mechanical aspects of the act of reading.

Structured and Unstructured Observation

In the early stages of the investigation of a given phenomenon, it is necessary to allow maximum flexibility in observation, for only through a flexible approach can a true picture of the phenomenon as a whole be obtained. Premature attempts to restrict the observation to areas considered significant entails the risk of overlooking some of the more crucial aspects. The investigator must be ready to shift from his original plans to the study of aspects which he sees as more significant. As the investigation proceeds, and the phenomenon is seen with greater clarity, the investigator must orient observation toward the more precise investigation of restricted aspects of the situation in an attempt to derive more rigorous generali-

zations—that is, he must orient his observation toward the systematic study of those aspects which previous research has shown to be significant. Eventually, these aspects can be subjected to even more precise investigation under experimental conditions.

Participant and Non-Participant Observation

When the subject of observation is the behavior of a human being—or perhaps an animal—the relationship that exists between the observed and the observer is of primary concern, since the very presence of the latter is likely to cause some shift in the behavior which he is trying to observe. Although the exact extent to which the observed is affected by the fact that he is being observed varies with such factors as the nature of the activity, the characteristics of the observer and of the individual observed, and so on, the reaction of the observed to the observer and to the observation itself is obviously a factor to be considered in connection with its validity. In fact, some research workers feel that, with the possible exception of observation through one-way screens and hidden microphones, it is not possible to observe without some distortion of the phenomenon being observed, and that all that can be done is to minimize such distortion and to take it into consideration in the interpretation of the results.[7]

Observation can be either *participant* or *non-participant.* In participant observation, the observer works his way into the group he is to observe so that, as a regular member, he is no longer regarded as an outsider against whom the group needs to guard. Sociological studies have been conducted in which the investigator joined groups of hoboes, hoodlums, and prisoners in jails in order to observe and understand them better. In non-participant observation, on the other hand, the observer remains aloof from the group. The fact that he is observing may be known to the group being observed, but the matter of his observation is made as inconspicuous as possible. In other studies, the observer may simply pretend to be a bystander, or

[7] Such distortion is, of course, not restricted to the social sciences: a similar phenomenon may be observed in physics, where for instance, it is realized that the apparatus used to detect the behavior of particles in atomic radiation distorts the movement of these particles—that is, the deflection of alpha particles is actually affected by the instruments used to measure them.

he may even hide behind one-way screens so that his presence is not even suspected.

The advantages and disadvantages of participant and non-participant observation depend largely on the situation. It is probably true that nothing can give a better insight into the life of hoboes, for instance, than living with them through inclement weather and other hardships. On the other hand, participation does not eliminate the distorting influence of the observer, for any member of a group must automatically play a role within that group. Furthermore, the participant observer is likely to adapt more and more to his role as a participating member of the group, and become more and more blinded to the peculiarities which he is supposed to observe. As a result, he is less likely to note what would be significant to a more objective observer. As he develops friendships with the members of the group, he is also likely to lose his neutrality and his objectivity and accuracy in rating things as they are.

Some research workers feel that it is best for the observer to remain only a partial participant and to maintain his status of scientific observer apart from the group. They claim that the distortion caused by the presence of the observer is not serious. It has been found that people, particularly children, get used to the observer to the point where they are no longer affected by the fact they are being observed. After a short period of adjustment, they simply resume their usual behavior. At the empirical level, studies have shown little difference between the observations made by observers out in the open and observers sitting behind one-way screens,[8] but this would vary with the nature of the phenomenon being observed and with other factors mentioned previously. It must also be remembered that certain observers are more capable than others of blending into a situation—either as participants or as external observers.

Recording Observation

An important aspect of observation concerns the recording of what is being observed. Specifically, the need for immediate recording, in order to minimize distortion due to forgetting, needs to be balanced against the inherent danger that the proc-

[8] R. F. Bales, *Interaction Process Analysis* (Cambridge: Addison-Wesley, 1950).

ess of recording will cause the observer to miss significant observations and will maximize distortion by making the fact of his observation conspicuous. The specific way in which the recording of observation can compromise with these two conflicting considerations varies from situation to situation. In certain circumstances, recording can be done directly, perhaps even with the help of cameras, sound tapes, and other mechanical means. This type of recording is ideal, since it gives an animated picture to be studied at leisure and capitalizes on the dynamic aspects of the situation. It also permits other observers to study the records and pass judgment on the adequacy of the interpretation.

In most instances, however, such ideal recording is out of the question. Taking longhand notes is generally inadvisable since it is too time-consuming and likely to cause impairment of the observational process. Shorthand may have advantages in certain cases. Probably the most commonly used, and generally the most practical, means of recording observational data is through the use of a checklist consisting of key words, which the observer can check as he goes along and from which he can later reconstruct the observation. The checklist is prepared in advance for the purpose of focusing the attention of the observer on relevant aspects of the situation and of systematizing his observation so that none of the significant aspects is overlooked. The checklist should be comprehensive and yet short enough to permit easy location of the items. The categories should be simple: a "yes," "no," or a key word to identify the alternatives is generally sufficient for an observer who is familiar with the situation.

There are times when any form of record-taking during an interview is inadvisable, and the observer must rely on his memory for the reconstruction of his observations. In such instances, he should record his observations as soon as possible after he leaves the setting, for the danger of distortion through forgetting is ever-present. There are, on the other hand, occasions when the appraisal that is required can be made only on the basis of the total observation, and any premature attempt at judgment will produce only incomplete and inaccurate appraisals which will bias later observations. In such instances, postponed recording can promote greater validity in observa-

tion by permitting the total picture to be seen in perspective. Of course, even in such cases, some kind of observational guide can be used to prevent the observer from overlooking any significant aspect of the situation.

Interpretation of Observations

Observation and the recording of observation are crucial steps in observational studies, for obviously any research technique must depend on reliable and accurate data. Even more important from a scientific point of view, however, is the interpretation of data from the standpoint of the problem under investigation. The observer in a research study is more than a machine merely registering what is going on; he is a scientist investigating a problem. And, while the interpretation of certain observational data is relatively obvious, in other instances drawing meaningful inferences from the data may require a high level of scientific sophistication and imagination.

Interpretation has to be done by someone somewhere along the line of investigation. It can be done directly by the investigator at the time of his observation. Favoring such an approach is the fact that the observer may be in a better position to interpret what he observes than someone who has to reconstruct the situation secondhand. On the other hand, the observer may have his hands full keeping up with what is going on, and any attempt at interpretation may distract him from his task of observation. Furthermore, where several observers are involved, on-the-spot interpretation introduces the problem of uniformity in interpretation. In such instances, it may be best for the observer merely to record his observations and to leave the matter of interpretation to an expert who is more likely to provide a unified frame of reference. It must, of course, be recognized that the interpreter's frame of reference is fundamental to any interpretation, and it might be advisable to insist on agreement between interpreters of somewhat different backgrounds and orientation as a means of counteracting possible bias in the results of a single observer.

Validity and Reliability of Observation

The validity and reliability of observational data are established on essentially the same bases as those for data derived

through other survey techniques. There are, however, a number of problems which are relatively peculiar to observation.

1. Establishing the validity of any appraisal is always difficult; it is especially difficult in observation, since many of the topics that lend themselves to an observational approach cannot be defined with sufficient precision to permit the isolation of their various aspects into different levels of relevance and significance. To attempt to define or to isolate these aspects may well involve false definitions and, consequently, invalidity. The problem of subjectivity is also involved in reconstructing the phenomenon through the addition of its component parts. Despite these inherent dangers, however, the derivation of carefully defined categories bearing directly on what is observed in the light of the purpose of the study is generally a prerequisite to valid and reliable observation. Care must be used, of course, not to concentrate on aspects of limited significance simply because they can be recorded objectively and accurately.

2. Inherent in observation is the possible distortion of the phenomenon through the very act of observing and the consequent introduction of bias, the direction and extent of which is relatively unknown and unknowable. Such distortion is difficult to eliminate, but it can be minimized through the proper choice and location of observers, inconspicuous recording, and other attempts at establishing observer neutrality.

3. A third difficulty peculiar to observation is that of obtaining an adequate sample of data on which to base conclusions. Since the observer has little control over the physical situation, it is frequently difficult to get information sufficiently free from complicating co-occurrences to give a clear picture of what is involved. This is particularly true in an unstructured situation where so many things can occur at once that it is difficult to attend to them all. The particular aspect of the situation in which the investigator is interested may occur so infrequently, and under such a variety of confounding circumstances, that it is difficult to establish its validity with any degree of precision.

4. The validity and reliability of observation depends primarily on the competence of the observer. Not only must he be fully trained in observational procedures and have a clear perspective of the nature of the phenomenon under study, but he must also have a valid frame of reference and a relative freedom from personal biases. Generally, greater validity and reliability is obtained by having two or more observers make paral-

lel observations. On the other hand, multiple observers are not a guarantee of validity, since they all may be subject to the same bias. If they have the same background and the same orientation, they are likely to look for the same things and to see and to interpret these observations in the same way.

Like that of the interview and the questionnaire, the empirical literature on the validity and reliability of observation is not very helpful, since validity and reliability depend on the specific nature of the variable in question and the conditions under which the observations take place. It is very difficult to generalize: what works in one situation may not work in another, and each situation has to be analyzed on its own merits. It seems reasonable to think, however, that, wherever it is properly used by competent observers under good research conditions, observation can yield results of scientific value and usefulness at all stages of the investigation of a phenomenon—from its early exploration to its final refinement. Furthermore, one must remember that observation is frequently the only means available for the investigation of certain phenomena, at least at their present stage of scientific development.

RATING STUDIES

Nature of Ratings

Many of the variables with which research is concerned cannot be measured directly; the degree of their existence has to be estimated on the basis of subjective judgment. In the social sciences, especially, the variables are frequently of such a nature that they can be ranked or rated only in crude categories along certain continua into which they can be ordered on the basis of their properties. Ratings are not restricted to the social sciences, however; phenomena which can be measured precisely—for example, musical pitch—are also frequently simply rated or ranked.

The concept of *rating* is probably best known in the area of tests and measurements, where—in its basic sense as a form of classification of items into levels along a given continuum—it parallels measurement. Rating differs from measurement in the refinement with which classification can be made and, especially, from the standpoint of the subjectivity involved. Measurement generally calls for nothing more than the skill to read

an instrument. Rating, on the other hand, implies the ability to estimate the status of a phenomenon or trait—that is, to make a subjective judgment of its status. As such, it differs from evaluation and appraisal which attempt to relate such status, measured or estimated, to adequacy from the standpoint of values, objectives, or other standards of reference.

The distinction is not quite so clear-cut, however. In rating a theme on a continuum of *excellent-poor,* for example, a value-judgment seems to be implicit in the scale values used. In other instances, the evaluation component, though not quite so clear, may be implied. Thus rating teachers on the basis of *democratic-autocratic* is not always divorced in the mind of the rater from the concept of desirability-undesirability.

As used in research, the term *rating* is generally given the more liberal meaning, and *rating studies* are generally considered essentially synonomous with *appraisal studies,* with considerable attention being devoted to the evaluation of what is discovered with reference to stated criteria of expectancy and desirability.

As a research technique, rating is a relatively crude and, as yet, undeveloped procedure, particularly with respect to the more significant aspects of education—attitudes, character, adjustment, leadership, values, and so on—the crucial aspects of which are still relatively undefined, and the tools for the appraisal of which are generally of limited adequacy. At present, such ratings incorporate a high level of subjectivity, and, while subjective judgment is involved in all research, there is need for restrictions to be placed on the extent and the manner in which judgment is allowed to influence the study. The advances that have been made in this area, while notable in the face of the complexities involved, are nevertheless relatively inadequate.

Mechanics of Rating

Ratings can be obtained through one of three major approaches: 1. *paired comparison,* 2. *ranking,* and 3. *rating scales.* A brief overview of each will be presented here simply as an orientation; the topic is too broad and complex to be covered adequately, and the student interested in such research should consult some of the references listed at the end of the chapter.

The first attempt at rating personality characteristics was the man-to-man technique devised during World War I. This technique calls for a panel of raters to rate every individual in comparison to a "standard person." Such an approach is feasible whenever the panel is well enough acquainted with each individual that they can make valid ratings. It could be used, for instance, in the rating of faculty members in a given department. A similar approach is to compare every single individual in a group with every other individual, and to arrange the judgments so derived in the form of a scale. Such techniques are, of course, extremely laborious when they involve a number of individuals and when they are extended to a number of variables. Another method used in the early stages of the development of rating techniques was to classify individuals in rank order with respect to a given trait, and to use the average of the rankings which a person received from a panel of raters as his personal scale value. Again, the procedure requires that the raters be rather thoroughly acquainted with all individuals to be rated.

The more common and more practical method of rating is based on the rating scale, a procedure which consists of assigning to each trait being rated whatever scale value seems a valid estimate of its status, and then combining the separate ratings into an overall score. The rating scale is best conceived as an instrument which permits the quantification of observation through the assignment of numerical values to the ratings of the various components of a given phenomenon, and the summation of these ratings into an overall index of its status. It is assumed that a more valid appraisal of a phenomenon can be obtained by the summation of the separate ratings of some of its critical components than by a general overall judgment.

The selection of the specific aspects of the phenomenon to be singled out for independent rating is based on their significance for the purpose of the investigation and their amenability to rating in the specific situation. Such decisions involve an element of subjectivity, and since the separate ratings are combined into a composite rating, failure to consider an important aspect of the overall trait immediately implies some degree of invalidity in the overall index. For example, whether

a rating scale for appraising the quality of handwriting should incorporate a separate rating of the firmness of the strokes may be subject to debate, but failure to incorporate the component of writing speed would tend to invalidate the rating, just as overemphasis on the factor of subject-matter competence might lead to an invalid rating of overall teaching ability.

Self-Ratings

In contrast to paired comparisons and rankings, which must be based on ratings by an outside rater, ratings on a rating scale can be made either by an individual or by an outside observer. Both approaches are commonly used and, of course, each has advantages and disadvantages. The question of self-observation and self-report has received the attention of psychologists since the beginning of psychology as a science. Originally considered under the term *introspection,* self-rating was the basic tool of discovery in the early days of psychology. With the shift of psychology toward behaviorism, however, anything that was not sufficiently overt to be verifiable by outside observers became suspect. At present a more lenient position has been taken toward self-observation and self-report on the obvious grounds that it is frequently the only means available for investigating certain relatively crucial aspects of the individual's psychological make-up.

Psychologists are fully aware of the limitations and potential dangers of self-reports. It is realized, for instance, that the minute a person becomes conscious of his reactions, he tends to change them so that they are no longer what he thinks they are. It is felt that self-reports are generally better measures of the person's self-concept than they are of the self in reality. Self-reports are predicated on the assumption that the individual understands himself—an assumption which psychologists would question, since individuals frequently have a very limited insight into their own dynamics. A prejudiced person, for example, does not see himself as prejudiced just as the humble person does not—and cannot—see himself as humble. The problem is further complicated by the reluctance of most individuals to reveal even what little they know about themselves. Not only are they likely to suppress the expression of what they consider self-depreciative, but they are also likely to

emphasize some of the positive aspects, beyond the element of truth. Unfortunately, even this is not universal; some people confuse the matter further by making themselves appear in the worst possible light. In other words, not only is the individual a poor judge of himself, he is also a biased reporter: his report tells us not what he is but what he feels (perhaps unconsciously) he is or would like us to believe he is.

These weaknesses, however, are not sufficient grounds for the absolute rejection of the self-report as a research technique. There is undoubtedly some degree of validity in the method, and it has been used with some degree of success even in such delicate areas as the appraisal of attitudes. Its limitations, however, must be clearly recognized, and investigators should be cautious in its use. The self-report is probably best used in the early stages of investigation as a means of providing hypotheses which can then be tested by more rigorous means. Its use would have to be evaluated on the basis of usual criteria of validity and reliability as they apply in the specific case. An extension of self-ratings which has considerable possibilities as a research technique in this area is the *Q-methodology* devised by Stephenson.[9] The procedure is beyond the scope of the present text and the reader is referred to the original source, or to other references listed at the end of the chapter.

External Ratings

The adequacy of external ratings also must be evaluated on the basis of the specific case. A primary limitation of their use is the basic question of whether the person (or thing) being rated is sufficiently well known to the rater for him to make a valid rating. It would also be necessary for the rater to have a perspective as to what constitutes *average, above average,* and *below average* status in the trait in question. A more valid rating is generally obtained, for instance, by pooling the ratings of a number of judges who have been carefully selected on the basis of their expertness with respect to the trait in question. It is also best to allow for *No information* in order to prevent uninformed ratings from vitiating the overall index.

A common error in rating is the *halo effect,* which may be

[9] William Stephenson, *The Study of Behavior: Q-Technique and Methodology* (Chicago: University of Chicago Press, 1953).

described as a general tendency for the rater to rate each of the individual's specific traits on the basis of a general overall impression or mental outlook, rather than on the basis of the traits as they appear independently. This is sometimes combined with the error of *central tendency* in which the rater, whenever he is not sure of the rating he should give an individual, rates him close to the average. Some raters are particularly reluctant to rate anyone at the extremes; this sometimes stems from the *logical error* which involves a lack of clarity of the trait being rated, and a consequent tendency to play it safe. Another common error is the *generosity* or *leniency error,* in which a rater tends to rate almost everyone above (or below) average. This error is of particular concern when multiple raters are used, since a lack of a common point of reference makes for non-comparability of the ratings of the various judges. This lack of a common point of reference is further complicated by shifts in the point of reference of an individual rater, who may rate leniently at one time and severely at another. To minimize this difficulty it sometimes helps to identify certain scale values as points of reference—for example, a C grade is performance typical of the average freshman—or to develop actual models—for example, an "A" theme, a "B" theme, and so on. Such *specimen* (*product*) scales are commonly used in the rating of handwriting.

The rater obviously plays a crucial role in the validity and reliability of the ratings. It is necessary, for example, to ensure that the raters are sufficiently familiar with the phenomenon being studied to see its components in perspective. Furthermore, it must be recognized that each rater brings to the situation his personal biases, which may distort his perceptions and interpretations in varying degrees. The rater must also have a clear idea of the point of reference which is to act as the benchmark in his ratings. Asking grade-school children to rate their teacher on a scale of *superior, above average, average, below average,* and *inferior,* for example, presupposes that they have a clear concept of what an "average" teacher is or does.

It must also be remembered that sizable individual differences in ability to rate exist; not only are certain individuals poor raters, but, to complicate matters further, some indi-

viduals tend to be much poorer raters with respect to certain phenomena or certain individuals than with respect to others. In the rating of individuals, for instance, a new dimension is introduced by the fact that the rater must be familiar with the person he is rating, but yet must not be so close to him emotionally that he loses his objectivity. This need for emotional detachment would, of course, hold as well for other phenomena toward which the rater has definite attitudes. It must further be realized that certain important phenomena, because of their nature, cannot be rated with precision, and that overemphasis on validity and reliability in the ratings frequently promotes the accurate rating of the trivial and the neglect of the significant.

In order to obtain relatively valid and reliable ratings, it is essential to clarify the nature of the phenomenon to be rated in the light of the objectives of the study. Ambiguity with respect to the aspects of the phenomenon that are to be included in the rating, for example, is likely to result in some degree of invalidity in the ratings that are made. Such a danger can be minimized by analyzing the phenomenon into its basic components, each to be rated separately, and defining these in operational terms. For instance, if the trait "teacher effectiveness" is broken down into fundamentals, such as "Is he or is he not tolerant of pupil mistakes?" which are probably considered only vaguely and nebulously when the rater's judgment is made on an overall basis, a more adequate rating is likely to result. It must be realized, however, that breaking down a variable into its components to be rated separately raises the question of the adequacy of the breakdown, on the one hand, and the validity of the synthesis of these components into an overall rating, on the other.

Structuring the situation in which the ratings are to be made can sometimes bring about greater uniformity in the ratings by providing a better basis for observation and a more restricted point of reference. On the other hand, it may promote invalidity in that it may make the situation artificial to the point where the phenomenon being rated is no longer that which exists in the natural situation. Practice sessions, in which a group of raters attempt to reconcile the differences in their ratings of a given phenomenon, are particularly effective

in clarifying the nature of the variable involved, in pointing out personal biases, and in calibrating the ratings to a common point of reference. Unless and until a relatively high degree of concordance in the ratings is obtained as a result of such practice sessions, there is no point in proceeding with the study.

The Rating Scale

The scale on the basis of which the ratings are to be made has received considerable attention in the psychological literature, and a number of specific rules can be given for its construction. It is necessary, for instance, that the wording of the items be clear and free from suggestion as to what the answers should be. The number of scale divisions to be used depends on the problem and the purpose of the study. For example, for some items, a five-point scale of *excellent, very good, good, fair, poor* may be better than a three-point scale, which gives the rater less freedom of operation. On the other hand, a scale should probably never extend beyond seven scale points, since the categories provided should have psychological existence and be within the possibilities of accuracy in estimation. The more the scale construction structures the situation, the greater the uniformity of ratings it is likely to promote. On the other hand, such a structuring increases the danger of overlooking certain possibilities that were not anticipated, and thus it places greater responsibility on the scale constructor and emphasizes the need for a pilot study to act as a basis for making the final adjustments on the scale.

The problems encountered in the construction of a rating instrument range from the relative simplicity of preparing a checklist to the extreme complexity of devising a more advanced rating scale required for the study of complex variables. In its most primitive stage, the checklist might call for the rating of but one or two factors, or it might consist of an aggregate of separate ratings of semi-independent aspects of a given situation or phenomenon. In the rating scale an attempt is made to give the instrument overall unity in line with its stated purpose. Thus, though the term is sometimes used loosely, a rating scale is generally a relatively elaborate and comprehensive instrument with the items arranged on a single continuum which provides an overall score or index. Since this index is ob-

tained through the combination of the items of the scale, it is essential to ensure that every significant aspect of the overall phenomenon is considered in its proper weighting, and that, conversely, nothing but components of the phenomenon is included.

The development of a scale to meet the above requirements entails a number of complexities beyond the scope of the present text; the reader is referred to sources more specifically devoted to the principles and techniques of scale construction. Briefly, what is required is that the items of the scale be *scalable*—that is, all the items must be on the same continuum allegedly measured by the scale. A number of techniques have been devised to appraise the scalability or internal consistency of the items of a rating scale with a view to removing non-scalable items whose inclusion in the overall rating, since they are apparently related to aspects other than that toward which the overall scale is oriented, would lower its validity. Thus in Guttman's scale analysis technique,[10] items are evaluated from the standpoint of *unidimensionality* by relating the individual's ratings on each of the items to his overall rating. For instance, on a scale of *honesty-dishonesty,* the item "Would you steal money from a friend?" might be non-scalable because it introduces a second dimension—loyalty—and thus makes it possible for a relatively dishonest person to rate the item at the "honest" end of the scale, not because of honesty but because of loyalty. Similar scaling techniques have been devised by Thurstone, Likert, and others.[11]

THE CRITICAL INCIDENT TECHNIQUE

A somewhat more recent development in analytical research is the critical incident technique, developed by Flana-

[10] Louis A. Guttman, "The Cornell Technique for Scale and Intensity Analysis," *Educational and Psychological Measurement,* 7 (Summer, 1947): 247–79.

[11] See such sources as Allen L. Edwards, *Techniques of Attitude Scale Construction* (New York: Appleton-Century-Crofts, 1957); Harold Gulliksen and Samuel Messick (eds.), *Psychological Scaling: Theory and Application.* (New York: Wiley, 1960); Orval H. Mowrer, "Q-Technique—Description, History and Critique," in Orval H. Mowrer (ed.), *Psychotherapy: Theory and Research.* (New York: Ronald Press, 1953), pp. 316–75; H. H. Remmers, *Introduction to Opinion and Attitude Measurement* (New York: Harper, 1954); Warren S. Torgerson, *Theory and Method of Scaling* (New York: Wiley, 1958); and Warren S. Torgerson, "Scaling and Test Theory," *Annual Review of Psychology,* 12 (1961): 55–70.

gan[12] during World War II. The technique is based on the premise that a more adequate rating of a phenomenon can be obtained through the separate appraisal of its individual aspects, but it goes further and postulates that a more valid appraisal of the different components can be obtained by rating each with reference to actual incidents that might characterize possession or lack of possession of the trait in question. It assumes that a better rating of a worker's efficiency, for example, can be obtained if what constitutes "worker efficiency" is defined in operational terms—that is, if it is related to specific and critical incidents of behavior which define the trait in question. "A good worker is prompt; a poor worker is tardy." "A good worker takes advice and suggestion with appreciation; a poor worker resents criticisms and suggestions."

The technique has been adapted to the study of certain phases of education. Ryans, for instance, incorporated it in the development of his teacher-behavior rating scale. First, he classifies teacher behavior into three major continua: 1. understanding, friendly *versus* aloof, egocentric, restricted teacher behavior; 2. responsible, businesslike, systematic *versus* evading, unplanned, slipshod teacher behavior; and 3. stimulating, imaginative, surgent, and/or enthusiastic *versus* dull, routine teacher behavior.[13] Each of these is analyzed further into such aspects as "apathetic, alert teacher behavior" in connection with (3) above, each of which, in turn, is to be rated on a seven-point scale. In order to identify the trait in question and, thus, ensure the adequacy of the ratings, however, critical incidents of apathetic and alert teacher behavior are listed in juxtaposition, as shown below:

Apathetic	1 2 3 4 5 6 7 N	*Alert*
Pupils were inattentive; showed evidence of wandering attention; indifferent to teacher.		Pupils responded eagerly, appeared anxious to recite and participate.

[12] John C. Flanagan, "The Critical Incident Technique," *Psychological Bulletin*, 51 (July, 1954): 327–58.

[13] The steps are listed in the order of the final product. In the actual derivation of his scale, Ryans started by collecting critical incidents which he then organized by means of factor analysis into teacher traits, and then into broad continua of teacher behavior.

Apathetic	1 2 3 4 5 6 7 N	*Alert*
Pupils were listless; spiritless.		Pupils watched teacher attentively when explanation was being made.
Pupils were restless.		Pupils worked concentratedly, appeared immersed in their work.
Pupils participated half-heartedly, assumed a "don't care attitude."		Pupils were prompt and ready to take part in activities.[14]

The analysis of the adequacy of the teacher, then, is not a matter of general impression but is related to dimensions of actual teacher behavior that are apparently crucial in spelling out the distinction between good and poor teachers. Similar studies could be made of democratic and autocratic leadership, for example.

FACTOR ANALYSIS

Another technique in this category which is widely used, particularly in the field of psychology, is *factor analysis.* Essentially, its purpose is to reduce a matrix of inter-correlations among test scores and other variables to a smaller number of psychological dimensions which will account for the wide diversity of individual performance. It attempts to analyze the inter-correlations among variables on the basis of broad factors, with a view to discovering the smallest number of such factors needed to account for the variance in the performance of the individuals represented. It can lead to the development of certain hypotheses concerning the relationships among these variables, which then can be tested experimentally. The purpose of factor analysis is not to test the significance nor to predict the occurrence of phenomena, however, but to analyze the *factorial composition* of a mass of data.

Factor analysis can provide valuable insights into the nature of phenomena, which then can be translated into a saving

[14] David G. Ryans, *Characteristics of Teachers* (Washington: American Council on Education, 1960).

of time and effort. For example, to the extent that factor analysis can show that two tests are measures of the same psychological trait, each is a duplicate of the other, and the use of both tests to measure this factor is unnecessary. Factor analysis can lead to a factorial purification of psychological tests and a consequent reduction in the degree of overlapping among them.[15]

As a technique of science, however, factor analysis is subject to a number of limitations. 1. It rests on the concept of correlation, with all of its inherent inaccuracies and weaknesses. 2. The factors have to be identified on the basis of judgment, and it is difficult to determine how many factors should be extracted from a given correlational matrix. 3. Only factors that have been included in the inter-correlational matrix can arise from factor analysis. Any factor that cannot be measured precisely with our present instruments is not likely to emerge as a factor, nor is a factor, such as ability to read, which is a common denominator in the matrix. There is no way of determining *the* factorial structure underlying human behavior: the only factors that can be discovered are those that have been included in the matrix. Consequently, the factorial pattern which is discovered in a particular study cannot be interpreted as conclusive evidence of the existence or significance of the factors discovered. There are also certain assumptions underlying the procedures that determine which factors will be extracted—that is, each solution is not unique but is dependent on the postulates that are accepted in the process of the derivation of the method. Thus whether we discover a general factor in the area of intelligence, as Spearman[16] did, or whether, like Thurstone,[17] we do not find a general factor depends on the basic assumptions from which the different techniques have been derived. That the same factors tend to emerge from dif-

[15] Scale analysis has revealed that probably all tests are factorially impure, many of them to an objectionable degree. Many of the items of the average test are not directly on the continuum indicated by the purpose of the test, but constitute vector forces whose net contribution to what is being measured is somewhat reduced by the fact that such contribution is made only vectorially. Not only does this lead to unnecessary test length for a given degree of precision, but it also makes for a certain degree of invalidity in the results.

[16] Charles E. Spearman, *The Abilities of Man: Their Nature and Measurement* (New York: Macmillan, 1927).

[17] See Chapter 15.

ferent studies simply reflects on the fact that when you start at the same starting point and proceed according to the same assumptions, you generally arrive at the same destination.

SCHOOL SURVEYS

Nature of School Surveys

A school survey generally is a comprehensive study of existing educational conditions, undertaken to determine the overall effectiveness of the school program with a view toward improvement where indicated. In a sense, it is a form of accounting or inventory. It gathers information about the various aspects of the school program and evaluates them in the light of the objectives of the school. It can be restricted to one specific element or one specific department, but, in general, it is most useful when it is designed to encompass the school program in its entirety.

Although the school survey is primarily directed toward the practical aspects of education rather than toward the development of education as a science, under proper leadership such a survey can lead to a scientific investigation of the causes of the weaknesses uncovered by the survey. The school survey can help clarify educational goals at the local level and reduce the gaps that exist between educational theory and educational practice. By forcing teachers to keep abreast of current developments, it helps to raise the standards of educational practice. School surveys vary in scope and complexity as well as in scientific sophistication, depending on the needs of the local situation and the capabilities of the personnel involved. The literature on the subject is voluminous and should be consulted for specific references. The various editions of the *Encyclopedia of Educational Research* have particularly good discussions of the nature and purpose of school surveys and their contribution to the cause of education.

Historical Development

The school survey is not new; well over a hundred years ago, Horace Mann and Henry Barnard were inspecting schools and making recommendations for their improvement. The first formal school survey was made in 1910 when the super-

intendent of Indianapolis was invited to make a survey of the schools of Boise, Idaho. This gradually became the pattern for early surveys: a neighboring superintendent, a professor from a nearby university, or, perhaps, an official of the United States Office of Education was invited to survey certain aspects of the program of a school system. These surveys were largely of an inspectional nature and frequently generated apprehension, as well as opposition, on the part of the local teachers, inasmuch as they generally ended with such a long list of needed improvements that they made local teachers feel quite insecure. Furthermore, such surveys generally lacked continuity from the standpoint of the implementation of the recommendations, and thus were of limited overall value.

The current trend is toward a more comprehensive study designed to evaluate the school as a functional unit. The usual survey begins with the clarification of the objectives of the particular school, and of education in general, and includes an appraisal of the administrative aspects, the instructional program, the physical plant, pupil transportation, personnel, pupil guidance, and so on—that is, every aspect of the school is considered in the light of these objectives. It is felt that, because of the interdependence of the various aspects, a survey of the overall program generally gives a more meaningful picture than a survey restricted to a single department in isolation. The trend is toward the developmental type of survey oriented toward making proposals for the improvement of the school, rather than toward the determination of existing conditions with emphasis on the discovery of weaknesses. It is common, furthermore, to attempt to maintain continuity from one survey to the next by orienting the attention of the investigating team to the strengths and weaknesses discovered in previous surveys and to the recommendations that were made. This generally acts as an incentive to implementing the recommendations of the previous evaluators and, thus, promotes a continuous program of self-improvement on the part of the school.

The school survey can be considered a case study of a school system—utilizing the results of survey testing, questionnaires and interviews, observation, ratings, and so on, and often enlisting the efforts of consultants and interested community leaders as well as local and neighboring school

personnel. The specific steps of the survey vary, of course, with the purpose and the scope, as well as with the caliber of the personnel involved. Generally, however, the major steps include: 1. the determination of the aims and the goals of the school; 2. a critical appraisal of the present program and its outcomes; and 3. an evaluation of the present operation from the standpoint of the objectives. It generally ends with recommendations for improvement.

Organization of School Surveys

The school survey can be conducted in one of three ways: 1. by outside consultants, 2. by the personnel of the local school system, or 3. by a community-wide group consisting of local teachers, interested members of the community, and resource persons, headed by a specialist acting as consultant and survey leader.

1. *The Consultant Survey.* Each approach has its advantages and its disadvantages. As we have noted, the early surveys were conducted by imported specialists, and, undoubtedly, a capable person can make such a survey meaningful and profitable. This method is, however, subject to certain limitations. Since a survey generally is of wide scope, a large number of weaknesses will probably be found, and the consultant is likely to find himself having to recite a long list of trouble spots. Anticipating this, teachers are likely to be reluctant to co-operate in the discovery of their weaknesses. Furthermore, since the teachers have almost no part in the survey, they are likely to be only mildly interested in implementing its recommendations; they may rationalize that the specialist did not have time to get a good picture—a point often well taken since, even with the full co-operation of the local personnel, it is difficult to get a good picture of the overall situation in the limited time available. Thus, too often the report is discarded the day it is filed, while teachers feel satisfied they have done their bit when they have agreed, or denied, that the problems exist or that nothing can be done about them.

This is not to say that specialists should never be used for making a local survey. Cornell lists the following situations as pointing to the need for an outside expert: 1. when the local personnel have been unable to cope with a problem; 2. when

the problems are so comprehensive that they create an overburden on the teachers, and 3. when there has been such inbreeding that there is a need for an external perspective.[18]

The consultant frequently has an advantage in that he can see the school in the light of his previous experience in similar schools. Not only is he more alert to the problems that may have become blind spots to those in the system, but he also can be more objective. The consultant usually has highly specialized training in research, and he frequently has a research design which can be implemented with minor modifications. Generally he has more prestige by virtue of being an outsider and frequently can command greater co-operation. On the other hand, while the expert in industry is a well-known figure, it must be recognized that the problems in education are so complex that perhaps no one can be an expert in all areas. It must also be fully recognized that if the expert is to be effective, everyone must give him full co-operation, which may be too much to expect when one realizes that people do not cooperate even with their own physicians. It is imperative, for instance, that faculty rating not be combined with the survey, if maximum improvement of the school program is to be attained.

2. *The Self-Survey.* The self-survey, involving teachers of the local school system working alone (except for perhaps a supervisor from the administration office) rarely accomplishes what a school survey should. Because they are close to their own problems, the teachers are likely to be blind to difficulties which have become so commonplace that they are no longer considered problems, or so ingrained that they are considered beyond solution. Furthermore, teachers frequently lack the insight into the true nature of their problems—and their potential solutions—which only the combined talents of the expert and of the teachers working as a team can provide. It is also true that teachers cannot shift gears easily; the reason their problems exist in the first place is that they have not had the competence to deal with them. Organizing a committee or a survey to give public testimony to the existence of the problem is not

[18] Francis G. Cornell, "Getting Action by Means of the School Survey," in *Growing Points in Educational Research.* (Washington: American Educational Research Association, Official Report, 1949).

much help. Furthermore, to the extent that the survey represents added responsibility imposed on their regular duties, teachers are more likely to resent the extra work than they are to look for solutions. Self-surveys are, of course, helpful when they are conducted in preparation for a more formal evaluation by a team of experts. In such instances, the teachers are more likely to be motivated to make improvements in preparation for the final survey, especially if they are encouraged to participate in and to make contributions to the latter.

3. *The Comprehensive Survey.* It is generally felt that the best approach to school evaluation is the comprehensive survey in which the school supplements its own personnel by enlisting the co-operation of teachers from nearby schools, interested community leaders, and consultants from neighboring universities and other school systems—all working together as a team under the direction of a steering committee headed by a survey leader. This broad approach is psychologically and administratively sound. Including the teachers in the survey—and in the improvement of their schools—is generally considered the best way of ensuring the success of the diagnostic aspects of the survey, as well as of promoting the implementation of the recommendations.

Participation of selected and interested community leaders is conducive to acquainting the community with its school, its philosophy, its operation, and the limitations under which it functions, and to promoting community goodwill. Laymen become less suspicious when they realize that the school has nothing to hide, that its problems are not insurmountable, and that its personnel is sincere in its desire to improve. The school needs to involve its patrons in identifying the problems which it faces. Community leaders can frequently bring a fresh approach to the problems of education that teachers have been too close to see. These are men of experience and often of wide and successful background who have much to contribute to the success of the school—not to avail ourselves of their services is shortsighted.

It is generally desirable to involve the public from the start by announcing the coming survey, by enlisting community help, and by keeping patrons informed through periodic reports. It is essential that the results be reported in such a way

that they will not be misinterpreted. This may mean a condensed report to be included without alterations in the local press, verbal reports through the local ETV channel, where possible, and perhaps brochures distributed to parents, civic clubs, and other community groups. Even when the study has been conducted by an outside expert, it is essential to report the results to the community. The report should be as clear as possible and oriented to the layman as well as to the educator. The emphasis should be on improvement and on the specific steps needed in such improvement.

4. *Surveys by Accrediting Agencies.* Surveys of schools are also conducted by accrediting agencies. While, at one time, such surveys were oriented toward the policing of higher education and the maintenance of standards, the present emphasis is on helping the school develop a worthwhile program in line with its objectives and plan for continuous improvement. It is invariably advisable for the school to be evaluated to conduct its own self-survey in preparation for the visit of the accrediting team. This will not only expedite the work of the accreditors but will also permit the school to get a feel of its strengths.

Each survey must be conducted in the light of the objectives of the school—and the community which it is to serve—rather than on the basis of an absolute standard, and though the evaluation involves the use of criteria or guides, their use is largely to orient the thinking and to prevent the overlooking of significant aspects rather than to set definite goals that need to be met. It is realized that two schools with identical programs are not necessarily equivalent and that inter-school comparisons are essentially meaningless, except in broad general terms. For this reason, it is no longer the practice of evaluators to give numerical ratings to the various items with a view to summating them in an index of overall quality. Again, the report should be precise in its recommendations concerning the specific ways in which weaknesses can be strengthened, but it must especially emphasize the strengths on which the school can capitalize in developing a program capable of growth.

Evaluation of School Surveys

School surveys have undoubtedly done a great deal to improve educational practice, especially when they have in-

volved local participation under competent leadership. Although no generalization can be made about the improvement which results from such a survey, there is no doubt that, if it is conducted under proper auspices, it can be effective. If maximum benefits are to be derived from such a study, however, it is essential that the evaluators avoid promoting undue uniformity and conformity; thus stifling the initiative, originality, and, particularly, the individuality a school needs to possess if it is to provide a program adapted to the needs of the community it serves. The survey should, for example, avoid placing undue emphasis on objective data on the basis of which to make inter-school comparisons, with a corresponding neglect of the more significant aspects which make a program functional, though different. Of even greater importance from the standpoint of the benefits to be derived from a school survey is the need for the school to provide for continuous appraisal of the extent to which the recommendations are implemented. This is probably best effected through the efforts of a research bureau in the central office working closely with the community and the local school.

SOCIAL SURVEYS

The community or social survey, while essentially the same from the standpoint of the research procedures involved, differs from the school survey in that it is more general and comprehensive and in that it is concerned with the school only as one of the many agencies whose function is vital to community welfare. Although it is not likely to probe so deeply and so specifically into the workings of the school as the school survey, the community survey is of great benefit to the school in clarifying the social setting in which it exists and functions, and the expectations of the community with respect to the education of its citizens.

Social surveys date back to such classic studies as that of slums and poverty in London by Charles Booth in the late 1800's,[19] and many others of a similar nature. More recent and more relevant to the work of the school in American society are

[19] Charles Booth, *Life and Labour of the People of London* (17 volumes; London: Macmillan, 1892–1897).

studies, such as that of Elmtown,[20] Yankee City,[21] and other communities, which have provided a considerable background of information of direct benefit to the school interested in serving its patrons more adequately.

Social surveys can be conducted by the local government, by local community leaders independently of the government, or by a team of experts imported by either of the above. They generally vary in effectiveness in proportion to such factors as the degree to which they stem from a clear-cut definition of a worthwhile problem, the care with which they are planned and executed, the resources and competence of the survey leaders, and the extent to which they enlist community participation.

The school should take an active part in the planning of such surveys in order to ensure that the factors bearing on its effective functioning are considered. The school would be vitally interested, for instance, in the educational status and aspirations of its citizenry, in population trends, and in the availability of cultural, religious, and recreational facilities and resources. The results should permit the school to see itself in perspective so that it can more effectively co-ordinate its efforts with those of the other community agencies.

GENETIC RESEARCH

Because of the importance of child growth and development to the whole process of education, *genetic* or *developmental research,* though not strictly an educational research technique, is of primary interest to teachers, particularly to those of the elementary school. Largely because of the obvious difficulties encountered in research of this kind—the time element, the extensiveness of the facilities required, the cost, and so on—this interest has been essentially from a consumer point of view. Except for studies of academic and perhaps intellectual growth—the growth in the child's ability to grasp concepts such as time sequence, for example—genetic studies are rarely selected for thesis or even dissertation pur-

[20] A. B. Hollingshead, *Elmtown's Youth: The Impact of Social Class on Youth* (New York: Wiley, 1949).

[21] W. L. Warner, *et al, Yankee City.* (Series; New Haven: Yale University Press, 1941–47).

poses. To date, the bulk of such research has been done in child-development clinics, among the most notable of which are those at Antioch College, Berkeley, Columbia, Chicago, Harvard, Iowa, Michigan, Minnesota, Stanford, and Yale. Among the many studies conducted in this area, those of Gesell and of Terman[22] are among the best known on the part of both professional personnel and the lay public.

Genetic research resembles a number of other research techniques described in this text. Genetic research, like historical research, is concerned with the occurrence of past events. It also approaches the experimental method, particularly when the development of identical twins is compared under slightly different environmental conditions.[23] It is closely related to survey methods in that it is concerned with the status of a phenomenon at successive stages of growth. It differs from all of these in purpose, however. It is not interested in the present status of development, nor in its historical background, nor even in the ways in which phenomena can be modified through the manipulation of environmental conditions—but simply in the pattern of development.

The techniques of genetic research have to be adapted to the age and nature of the subjects. For instance, in studying infants and pre-school children, it may be necessary to use direct measurements, observations through one-way screens, and so on. For older children, on the other hand, tests of the pencil-and-paper variety might be used. It can also vary in duration. For example, genetic studies of a short duration could be conducted with respect to such factors as academic growth, which is relatively rapid and for the measurement of which we have relatively adequate instruments. On the other hand, short-term genetic studies would not be effective for studying some of the more slowly developing aspects of growth, such as personality.

Genetic studies can be either *longitudinal* or *cross-sectional.* Longitudinal studies follow the same group of subjects over a relatively long period of time. For example, in 1958, Terman conducted his fourth follow-up of the one thousand

[22] See Chapter 15.

[23] See Arnold S. Gesell and Helen Thompson, "Learning and Growth in Identical Infant Twins," *Genetic Psychological Monographs,* 6 (1929): 5–120; and Myrtle B. McGraw, *A Study of Johnny and Jimmy* (New York: Appleton-Century-Crofts, 1935).

gifted children whose study he began in 1925.[24] A cross-sectional approach, on the other hand, consists of taking random samples of children of successive ages as the basis for developing growth norms. The longitudinal approach is generally considered more acceptable than the cross-sectional because it has the advantage of continuity and permits the recording of individual fluctuations, which are frequently of greater interest than the overall growth pattern itself. The time factor poses a special problem, however, especially to doctoral or master's students. The maintenance of co-operation on the part of the subjects and the loss of subjects over long periods of time also present difficulties. The cross-sectional approach, on the other hand, is particularly vulnerable to the sampling problem, so that a fairly large sample would have to be used at each of the successive age levels in order to provide valid data. It is sometimes possible to combine the two approaches by having, for example, four overlapping groups at two-year intervals. In this way, one might conduct in two years a study that would normally take eight, and, at the same time, validate each sample, one to the other, at the point of overlap. The time problem can also be overcome by conducting genetic studies through records, if adequate records have been kept. This condition is rarely fulfilled, however, unless careful plans have been made in advance. For example, it is likely that even such simple matters as the IQ are not recorded on a comparable basis over the years and, therefore, the required continuity is lacking.

The major weakness of genetic studies is that they give growth patterns that represent the average of the group and apply, therefore, only indirectly to the individual case. In physical-growth curves, for instance, there is no place where the pre-adolescent growth spurt is shown, simply because it is neutralized from one person to the other—with the result that the overall pattern is, in a sense, erroneous and misleading. Another weakness is that, though some attempt at theory construction in the science of development has been made, the approach so far has been essentially empirical in nature. The information it provides is, of course, useful, since it helps us to understand both the typical and the atypical child. Furthermore such an empirical approach is necessary in the beginning

[24] See Chapter 15.

stages of research. There is, however, the need to move on toward a science of growth and development. Science is concerned with the relationship among phenomena rather than simply with their existence; to note the status over the years and to develop a set of norms may be valuable, but it is only a preliminary step in the development of science which would be more concerned with the prediction and the control of the growth pattern.

SUMMARY

1. Analysis is primarily a fundamental technique of science, underlying all scientific procedures. The analysis of a phenomenon into its components permits the identification of its crucial aspects and provides a deeper insight into its nature and a more adequate basis for its allocation into meaningful classifications. The point in the analysis which provides the greatest insight varies with the nature of the phenomenon and especially, with the purpose of the investigation.

2. Analysis is also a research method, comprising a variety of techniques designed to dissect phenomena into their constituents as a means of providing greater insight into their basic nature. One of the most elementary of these techniques is documentary-frequency research.

3. Besides underlying all research, observation is also a research method in its own right and many phenomena can be investigated in no other way. Observation, especially unstructured observation, is probably the most flexible research technique and is consequently particularly suited to the early exploration of a given problem. However, because of its extreme flexibility and of the nature of the problems for which it is suited, it is frequently difficult to have observation meet the criteria of objectivity, reliability, and validity, required of a scientific data-gathering instrument. Scientific observation must be distinguished from the capricious and haphazard observation of the layman.

4. Just as in the case of the interview, it is particularly difficult at times to prevent the very presence of the observer from vitiating the observation, and again, the selection and training of the observers is a crucial aspect of its success. The recording of the observations should generally be done inconspicuously and as expeditiously as possible to minimize the danger of distorting the observation.

5. Observation is frequently quantified through rating, a procedure consisting of assigning numerical values to represent the various degrees of the phenomenon in question. Ratings—and especially ratings of some of the more important phenomena of

concern to educators and psychologists—unavoidably incorporate a high level of subjectivity and imprecision. Among the more common errors are the halo effect, the error of central tendency, the logical error, and the generosity error. Rating as an analytic technique is based on the premise that a more adequate appraisal of an overall phenomenon can be obtained by the separate rating of its various components, and that these separate ratings can then be combined into a meaningful overall index. The critical incident technique attempts to promote greater validity in rating by orienting the rater's attention to specific instances of the phenomenon in question.

6. Factor analysis attempts to telescope a vast array of different observations into a small number of underlying dimensions. Although factor analysis as a scientific technique has a number of limitations that must be clearly recognized, it serves a useful purpose in the clarification of phenomena.

7. School surveys constitute a form of inventory of the operation of the school considered in the light of its objectives. Although good results can sometimes be obtained through a survey conducted by an outside expert or by the personnel of the school, generally the most fruitful approach to school evaluation is the comprehensive survey which involves teachers from nearby schools, interested community leaders, and outside consultants working with and through the local school personnel. No matter what the approach, the emphasis should be on building up strengths rather than on identifying weaknesses, and plans must be instituted for the implementation of a program of continuous self-evaluation and self-improvement.

8. The community survey, while not so specifically oriented toward the operation of the school as the school survey, can provide valuable perspective into the function of the school in the community it is to serve.

9. Genetic research is interested in the pattern of growth over a period of time rather than simply in the measurement of its status at a given moment. Unfortunately, it is oriented toward the derivation of group norms and therefore it tends to bury individual variations in growth, which are frequently the most significant aspect of the situation. Despite inherent difficulties with respect to attrition of the research population and the time element, the longitudinal approach to genetic research is generally superior to the cross-sectional. To date, genetic research has been essentially empirical; there is need for a greater emphasis on the development of a theoretical orientation.

PROJECTS and QUESTIONS

1. Make a content analysis of a recent textbook in educational research. Develop evaluative criteria and rate its various com-

ponents to obtain an overall index of adequacy. Learn how to use a readability formula.

2. Plan for the classroom observation of both teacher and pupil behavior and note inter-observer agreement. Reconcile disagreements through discussion as a means of promoting validity, of gaining insight into the variables observed, and of getting a feel of observation as a research technique. Identify some of the characteristics of good and poor observers.
3. Prepare a score card for evaluating the various research methods discussed in this text.
4. Prepare a rating scale for appraising the emotional climate of a classroom.
5. Evaluate the report of a school survey in the light of its primary purpose of promoting improvement in the operation of the school. (Actual participation in an evaluation would be a most profitable experience.)
6. Describe the facilities of a psycho-educational clinic—such as that at Yale—for the conduct of genetic research.

SELECTED REFERENCES

BARR, ARVIL S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," *Journal of Experimental Education,* 16 (June 1948): 203–83.

BIDDLE, RICHARD A. "The Construction of a Personality Inventory," *Journal of Educational Research,* 41 (January 1948): 366–77.

BLOCK, JACK. "The Difference between Q and R," *Psychological Review,* 62 (September 1955): 356–8.

BURNS, HOBERT W. "Success Criteria and the C. I. Technique," *Phi Delta Kappan,* 38 (November 1956): 73–5, 256.

———. "The Critical Incident Technique as an Instrument of Educational Research: A Philosophical Analysis," *Dissertation Abstracts,* 18 (February 1958): 520.

CATTELL, RAYMOND B. "On the Disuse and Misuse of P, Q, Qs, and O Techniques in Clinical Psychology," *Journal of Clinical Psychology,* 7 (July 1951): 203–14.

CHALL, JEANNE S. *Readability: An Appraisal of Research and Application.* Columbus: Ohio State University, 1958.

COOMBS, CLYDE H. "Theory and Methods of Social Measurement," in Festinger, Leon and Katz, Daniel (ed.). *Research Methods in the Behavioral Sciences.* New York: Dryden, 1953.

———. *A Theory of Psychological Scaling.* Engineering Research Institute Bulletin, No. 34. Ann Arbor: University of Michigan Press, 1951.

COOPER, BERNICE L. "The Critical Requirements for the Principalship Based upon an Analysis of Critical Incidents Reported by Instructional Supervisors and Visiting Teachers," *Dissertation Abstracts,* 17 (February 1957): 290.

CORBALLY, JOHN E. "Critical Incident Technique and Educational Research," *Educational Research Bulletin,* 35 (March 1956): 57–62.

———. "A Second Look at the C. I. Technique," *Phi Delta Kappan,* 38 (January 1957): 141–2.

CORNELL, ETHEL L. and ARMSTRONG, CHARLES M. "Forms of Mental Growth Patterns Revealed by Reanalysis of the Harvard Growth Data," *Child Development,* 26 (September 1955): 169–204.

CRONBACH, LEE J. and GLESER, GOLDINE C. "A Review of Stephenson's: The Study of Behavior: Q-Technique and Its Methodology," *Psychometrika,* 19 (December 1954): 327–30.

DEARBORN, WALTER F. and ROTHNEY, JOHN W. M. *Predicting the Child's Development.* Cambridge: Sci-Art Publishing Co., 1941.

EDWARDS, ALLEN L. and KENNEY, KATHERINE C. "A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction," *Journal of Applied Psychology,* 30 (February 1946): 72–83.

FLANAGAN, JOHN C. "Research Techniques for Developing Educational Objectives," *Educational Record,* 28 (April 1947): 139–48.

———. "The Critical Requirements Approach to Educational Objectives," *School and Society,* 71 (March 27, 1950): 321–4.

———. *The Critical Incident Technique in the Study of Individuals.* Washington, D.C.: American Council on Education, 1952.

GESELL, ARNOLD S., *et al. The First Five Years of Life.* New York: Harper, 1940.

———. *The Child from Five to Ten.* New York: Harper, 1946.

———. *Youth: The Years from Ten to Sixteen.* New York: Harper, 1956.

GUILFORD, J. PAUL. "When Not to Factor Analyze," *Psychological Bulletin,* 49 (January 1952): 26–37.

GUTTMAN, LOUIS A. "The Cornell Technique for Scale and Intensity Analysis," *Educational and Psychological Measurement,* 7 (Summer 1947): 247–79.

JOHNSON, DONOVAN A. "Readability of Mathematics Books," *Mathematics Teacher,* 50 (February 1957): 105–10.

KERLINGER, FRED N. "The Attitude Structure of the Individual: A Q-Sort of the Educational Attitudes of Professors and Laymen," *Genetic Psychological Monographs,* 53 (May 1956): 283–329.

KING, WILLIAM B. "Certain Critical Requirements for the Secondary School Counselor Determined from an Analysis of Critical Incidents Reported by Teachers," *Dissertation Abstracts,* 19 (July 1959): 1946.

KROGMAN, WILTON M. *The Physical Growth of Children: An Ap-*

praisal of Studies, 1950–55. Lafayette: Child Development Publications, 1956.

LIKERT, RENSIS. "A Technique for the Measurement of Attitudes," *Archives of Psychology*, No. 140, 1932.

LORD, FREDERICK M. (Chairman). "Problems and Procedures in Profile Analysis," in *1953 Invitational Conference on Testing Problems*. Princeton: Educational Testing Service, 1953.

——— "Some Relations between Guttman's Principal Components of Scale Analysis and Other Psychometric Theory," *Psychometrika*, 23 (December 1958): 291–6.

MCNEMAR, QUINN. "Opinion-Attitude Methodology," *Psychological Bulletin*, 43 (July 1946): 289–374.

MAYHEW, LEWIS B. "The Critical Incident Technique in Educational Evaluation," *Journal of Educational Research*, 49 (April 1956): 591–8.

MEEHL, PAUL E. "Configurational Scoring," *Journal of Consulting Psychology*, 14 (June 1950): 165–71.

MESSICK, SAMUEL J. "Some Recent Theoretical Developments in Multi-Dimensional Scaling," *Educational and Psychological Measurement*, 16 (Spring 1956): 82–100.

NEW YORK BOARD OF HIGHER EDUCATION. *Studies of Teacher Behavior: 1. The Refinement of Two Techniques of Observing Teachers' Classroom Behavior; 2. Inferring Classroom Behavior from Pupil Responses*. New York: The Board, 1955, 1956.

NIVEN, JAROLD R. "A Comparison of Two Attitude Scaling Techniques," *Educational and Psychological Measurement*, 13 (Spring 1953): 65–76.

PISARO, SAMUEL E. "A Study of the Critical Elements in Teacher Effectiveness in Indiana City, Town, and Consolidated Schools," *Dissertation Abstracts*, 19 (August 1959): 1975–6.

PRESSEY, SIDNEY L. and KUHLEN, R. G. *Psychological Development through the Life Span*, New York: Harper, 1957.

RYANS, DAVID G. *Characteristics of Teachers*. Washington, D.C.: American Council on Education, 1960.

——— "Investigations of Teacher Characteristics," *Educational Record*, 34 (October 1953): 371–96.

SEEGER, RUTH C. *Readability—A Bibliography*. Columbus: Ohio State University, 1952.

SELLS, SAUL B. and ELLIS, ROBERT W. "Observational Procedures Used in Research," *Review of Educational Research*, 21 (December 1951): 432–49.

TERMAN, LEWIS M. *et al. Genetic Studies of Genius*. Vols. 1–5. Stanford: Stanford University Press, 1925, 1926, 1930, 1947, and 1959.

TURNER, RICHARD L. and FATTU, NICHOLAS A. "Skill in Teaching: A Reappraisal of the Concepts and Strategies in Teacher Ef-

fectiveness Research," *Bulletin of the School of Education* (Indiana University), 36, No. 3, 1960.

WATTERS, WILLIAM A. "Annotated Bibliography of Publications Related to Teacher Evaluation," *Journal of Experimental Education,* 22 (June 1954): 351–67.

WITHALL, JOHN G. "The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms," *Journal of Experimental Education,* 17 (March 1949): 347–61.

If we are to advance beyond the dark ages of educational pre-science, we must emulate the experimental proficiency and zeal of colleagues in other behavioral sciences.

JULIAN C. STANLEY

12 The Experimental Method

The Nature of Experimentation

Undoubtedly, experimentation is the most scientifically sophisticated research method. In fact, experimentation is frequently confused with the scientific method by the layman, who equates experimentation with the physical sciences and, further, equates the physical sciences with science itself. This erroneous viewpoint probably stems from the fact that the first steps in setting up experimentation as the ultimate in research were taken in the physical sciences. Actually, despite its scientific rigor, experimentation is only one aspect of the scientific method, for the scientific method involves a great number of activities of which experimentation is simply an important form.

The first "experiment" was apparently conducted by Galileo who, in 1589, showed that bodies of the same substance fall at identical rates of speed regardless of their mass. Early studies also were conducted in biology—for example, Pasteur's discovery that food spoilage can be attributed to bacteria. Similar experiments are common today in the field of medicine where the testing of drugs and vaccines is generally based on a relatively simple experimental design. Experimentation is

somewhat less of a standard procedure in the social sciences where the problems are generally complex. At the present stage of development of educational science, many of the more significant educational problems are not particularly amenable to rigorous experimentation—especially experimentation based on the simple experimental design discussed above, which is essentially inadequate for dealing with the complex problems with which education is concerned. Only since the development of *multivariate analysis* has experimentation into realistic educational problems become possible.

The purpose of experimentation is to derive verified functional relationships among phenomena under controlled conditions or, more simply, to identify the conditions underlying the occurrence of a given phenomenon. From an operational point of view, it is a matter of varying the *independent* variable in order to study the effect of such variation on the *dependent* variable. For example, the investigator might vary the size of the print and appraise the effect of such manipulation on reading speed. Actually, what we know about our environment comes from observation, and all research is concerned with the observation of phenomena and the generalization of these observations into certain functional relationships whose validity can be tested. Experimentation simply enables us to improve the conditions under which we observe and, thus, to arrive at more precise results. This is the essence of the scientific method.

The Concept of Causation

The concept of *causation* is always troublesome; it is troublesome even in experimentation where it is most fundamental. In the earlier use of the term, as illustrated by Mill's canons, causation implied an invariant one-to-one relationship between certain antecedents and certain consequents. In keeping with the resulting emphasis on the law of the single variable, the investigator attempted to control all relevant factors except the experimental factor, and thus he promoted certain outcomes which were then measured and attributed to the operation of the variable under investigation. Unfortunately, such ideal conditions are rarely fulfilled—even in the physical sciences.

In education, it obviously is impossible to equate two situations in all respects except for the factor whose effect is being investigated. In practice, it is not essential that the two situations be identical in every respect because many of the aspects in which they differ are irrelevant to the investigation and, therefore, can be ignored. However, this presupposes that one knows what is relevant and what is irrelevant, and obviously if a variable is assumed to be irrelevant when, in reality, it is relevant, some degree of error is introduced. In fact, failure to control all relevant factors except the one under investigation, either because of failure to see their relevance or inability to neutralize their influence, undoubtedly constitutes the prime source of erroneous conclusions derived from experimentation.

Quite apart from the practical limitations of experimentation based on the *law of the single variable* is the even more damaging objection that such control is theoretically unsound, since imposing control on a situation tends to make the situation artificial and the results meaningless. Thus, even if complete control could be obtained, it would only serve to violate the basic principle of maintaining a natural situation. This is particularly true in the social sciences, where the problem situations are invariably so complex that attempting to reduce them to the operation of a single variable simply defeats the purpose of the experiment by seeking a partial answer out of the context of reality. In fact, the unwarranted transfer of the law of the single variable from the physical sciences—where it might conceivably be used—to education, where it is essentially inappropriate, is responsible to a large degree for the relative unproductivity of educational experimentation to date.

This earlier interpretation of causation is unnecessarily narrow and mechanistic. Science is not interested in the effect of a single factor examined in isolation, but rather in the joint effect and interaction of many factors operating simultaneously. Furthermore, rather than an invariant one-to-one relationship between antecedent and consequent, the more realistic concept of causation is predictability—that is, the statistical probability of the occurrence of a given phenomenon in response to a given set of antecedents. Basic to this modern view are the concepts of *multiple causation* and *concomitance,* both of

which are especially fundamental to experimentation in the social sciences. The present interpretation is that experimentation must operate in the context of the complex *multivariate* interaction which characterizes phenomena as they actually exist.

Manipulation of Variables

Although the distinction is not always clear, experimentation differs from other methods of research largely from the standpoint of whether observation takes place under natural conditions or under conditions that have been deliberately staged to bring about the operation of a given factor and its resulting outcomes. Thus in contrast to the survey method, in which the investigator simply observes phenomena as they occur naturally, experimentation actually sets the stage for the occurrence of the factor whose performance is to be investigated under conditions in which all other factors which might confuse or complicate the observation are isolated or controlled. It permits the more rigorous allocation of the occurrence of a given phenomenon to the operation of a given factor—that is, it permits the more rigorous identification of the relationships among phenomena.

Experimentation is both economical and precise. Rather than waiting for a phenomenon to occur—and, of course, occur sufficiently often under such a variety of conditions that the irrelevant factors can be eliminated—the experimenter sets up the conditions that bring about its occurrence under the conditions most favorable for observation. Since these conditions can be varied by degree, the progressive effect which such manipulation has on the dependent variable can be evaluated. Making phenomena occur under specified conditions at a time when the investigator is ready to observe not only permits him to obtain more accurate answers, but also permits other investigators to verify his findings. Thus, while all science depends on observation, experimentation permits the investigator to increase the precision of his observations through controlling the conditions under which they take place.

The fact that experimentation takes place in a prearranged setting does not imply that the experimental situation has to be created *in toto*. On the contrary, the experimenter frequently

takes advantage of situations that already exist. For example, in investigating the effect of organic brain damage on test response, he does not start with identically normal individuals and subject some of them to varying degrees of brain damage in order to note the degree of impairment in mental functions. Rather, he capitalizes on groups as they already exist. Similarly, in the investigation of the effect of intelligence on academic performance, the comparison is made on the basis of pre-existing groups of children of different IQ's. Selecting cases with pre-existing characteristics—besides being more realistic—generally accomplishes the same purpose as does causing the characteristic to occur in order to note its effect.[1]

The problem of the manipulation of experimental variables has, of course, been simplified a hundredfold by the advent of modern statistical techniques, which allow groups to be equated statistically on certain variables on which complete equivalence cannot be readily effected through physical means. More recent developments in multivariate analysis have reduced the need for the physical manipulation of variables so that progressively more complex variables can be investigated with a minimum of alteration of the natural setting in which they exist. Such procedures, instead of relying on the neutralizing of influencing factors so that their effect on the phenomenon in question is eliminated, actually incorporate these factors into the design and isolate their effects through statistical procedures. They provide separate measures of the significance not only of the main effects of each of the variables but also of the interaction among them—the latter frequently being the most significant aspect of the overall situation. In a study of the effects of IQ, general scholarship, and grade placement on knowledge of current events among high-school students, for instance, it might be possible to obtain separate tests of the effects of the three factors taken singly, of the factors taken two at a time, as well as of the overall effect of the three factors taken all at once.

Multivariate analysis saves time and effort in that it permits the simultaneous investigation of a number of variables

[1] On the other hand, care must always be exercised in using pre-existing groups for research purpose, inasmuch as the investigator generally has no control over the complicating circumstances that may have been responsible for the groups being what they are. See the discussion on the *ex-post-facto* experiment on page 334.

considered singly and in interaction, and, up to a point, because of its ability to provide a more accurate estimate of the error involved, it does so with a relatively greater degree of precision than the simpler experimental designs. On the other hand, multivariate analysis calls for some background in advanced statistics and is, of course, beyond the scope of a text in introductory research. For a more complete orientation the student is referred to any one of the many textbooks in advanced statistics to be found in the library.

Experimentation is a more refined and advanced research procedure than most other forms of research. In the sequence of the investigation of a problem, one can begin with casual observation, unstructured interviews, and open questionnaires. As the nature of the problem becomes clearer, he might turn to more rigorous survey techniques, such as the structured interview, the closed questionnaire, and the various methods of analysis, in order to identify a complex of significant factors and to formulate more definite hypotheses regarding the possible relationships involved. The final step would be an experimental test of the hypotheses so derived. Thus, in contrast to the other methods of research, which tend to be more appropriate for the exploratory stages of investigation, experimentation attempts to provide a precise answer to a precise question. Conversely, experimentation cannot be used effectively until the area of investigation has been sufficiently defined—through some of the less rigorous, and, therefore, more flexible techniques—to permit the identification of the factors that need to be controlled, and the specific hypotheses that need to be tested.

CHARACTERISTICS OF AN EXPERIMENT

The purpose of experimentation is to identify functional relationships among phenomena through staging the occurrence of certain outcomes under controlled conditions designed to prevent the confusing effects of the operation of extraneous factors. Experimentation can be considered a technique of deliberately staging a situation designed to force nature to provide a "yes" or "no" answer to a specific hypothesis concerning the phenomenon under discussion.

If experimentation is to provide a meaningful solution to a problem, it is essential that the experiment contain, within itself, the means for answering its own questions—that is, the experiment must be self-contained. This, in turn, calls for the satisfaction of three basic and interrelated conditions—*control, randomization,* and *replication.* Unless these conditions are fulfilled, the experiment cannot be interpreted, for it cannot eliminate the possibility that the results obtained were caused by factors other than that under investigation. More specifically, the experiment must provide the basis for calculating the probability that the phenomenon which did occur was the result of the experimental factor rather than of the operation of extraneous factors.

Control

The basic element of experimentation is *control.* The experiment must be organized so that the influence of extraneous factors that are not included in the hypothesis are prevented from operating and confusing the outcome which is to be appraised. To illustrate: Assume that eight out of ten rabbits inoculated with Serum X are dead within twenty-four hours. The results are incapable of interpretation since the "experiment" does not exclude the possible influence of extraneous factors. The rabbits may have died as a consequence of the fright attending their capture to be inoculated, for example, or perhaps as a reaction to the disinfectant in which the needles had been lying. As Wilson points out, if one doubts the necessity of controls, all he has to do is to reflect on the statement: "It has been conclusively demonstrated by hundreds of experiments that the beating of tom-toms will restore the sun after an eclipse."[2]

For the above experiment to be self-contained, it would be necessary to have a *control group* of rabbits that parallels the *experimental group* in all respects, except for the fact that one group is inoculated with the serum and the other group is not. This simple design can be extended to include a third group acting either as a second experimental group or a second control group, depending on the interpretation this third

[2] E. Bright Wilson, *Introduction to Scientific Research* (New York: McGraw-Hill, 1955), p. 41.

group is to be given from the standpoint of the purpose of the study. In the study of the Salk vaccine, for instance, an experimental group of nearly a half-million children were inoculated, nearly a quarter-million were given placebos of distilled water, and over a million received no vaccine at all; the latter two groups acted as control for the experiment.

The control of relevant factors is, of course, very difficult to establish, expecially in the social sciences. This is particularly true in education where one does not violate the principles of good teaching just to carry out an experiment. Furthermore, the number of variables to be controlled—chronological age, intelligence, previous background, enthusiasm and motivation, study habits, the time available for study, the amount of outside work, and so on—is large. There is a limit to the extent to which one can manipulate human beings for experimental purposes, and, as far as educational research is concerned, it is frequently necessary to compromise between what is administratively feasible and what is scientifically rigorous.

1. *Danger of Artificiality.* Although control is fundamental to experimentation, care must be taken not to control the situation so that it becomes artificial and so that, consequently, the results, even though highly rigorous, are inapplicable and meaningless from the standpoint of the actual situation. The problem is well presented by Page,[3] who discusses the dilemma between controlling the situation so that while rigorous generalizations are reached they do not apply to any real situation, and investigating the situation in actual existence, thus permitting the application of the results to a similar situation but precluding any generalization. The pure scientist would insist upon generalizability, even though the control of conditions necessary to bring it about makes the situation artificial and its conclusions inapplicable. He would maintain that unless control is exercised the results are meaningless, since there would be no way of knowing what "caused" them. To him, generalizability is a more basic concept and replicability has to be sacrificed. The practitioner, on the other hand, would insist on applicability to the real situation, regardless of whether the results can be generalized to any general class of

[3] Ellis B. Page, "Educational Research: Replicable or Generalizable," *Phi Delta Kappan*, 39 (March, 1958): 302–4.

events. He would contend that rigorous results that apply nowhere are automatically useless.

This is the basic issue of *action research,* in which scientific rigor has frequently been sacrificed in order to obtain an answer to a practical problem existing here and now. To the extent that the investigator exercises rigorous control over a situation, he automatically establishes conditions different from those of the regular situation and by that very process alters his problem, and makes it impossible to apply his findings to an actual situation. This problem is illustrated by animal experimentation where, because of the high degree of control possible, the results are highly generalizable but hardly applicable to the human situation. Unfortunately, it seems that in the social sciences—where phenomena are invariably complex—what is discovered most precisely is very frequently what is least useful because it is most artificial.

2. *Control in Educational Experiments.* In classroom-teaching experiments, it is particularly difficult to control the enthusiasm and the zeal of the teacher and the motivation which he generates in his students. Almost any procedure—no matter how unsound—that stirs the enthusiasm of the children and their teachers is likely to be more effective than another method in which motivation is not at such a high pitch. To the extent that enthusiasm is frequently based on such transient factors as the novelty of the method, the findings concerning the true worth of the methods being compared are invalid. There would be nothing wrong, of course, with incorporating enthusiasm in the experiment when differences in enthusiasm are inherent in the methods. To the extent that the pupil-directed approach to learning is more closely synchronized with the child's needs, goals, and purposes, for example, it might be expected to have greater pupil motivation than the teacher-directed method. It would be incorrect to attempt to equalize pupil motivation in such a study, for it would destroy the variables under investigation.

Another factor that needs to be considered very closely is the nature of the experimental design itself. Occasionally, the experiment is designed in such a way that it can lead to only one conclusion. Such a situation is described by Bertrand Russell regarding the classic studies in the psychology of learning:

> One may say broadly that all animals that have been carefully observed have behaved so as to confirm the philosophy in which the observer believed before his observations began. Nay, more, they have all displayed the national characteristics of the observer. Animals studied by Americans rush about frantically, with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think, and at last evolve the solution out of their inner consciousness. To the plain man, such as the present writer, this situation is discouraging. I observe, however, that the type of problem which a man naturally sets to an animal depends upon his own philosophy, and that this probably accounts for the differences in the results.[4]

Frequently, the criterion against which the outcomes of a given experiment are measured is directly related to the outcomes the experimental method was designed to produce—so that the experiment is a success from that standpoint—but nothing is said about the price that was paid because of the neglect of other equally desirable objectives. It seems logical that any program designed to emphasize one phase of the overall academic program is likely to be successful in that phase of it. To the extent that the criterion emphasizes one phase of the overall program and minimizes others, the results are likely to favor one or the other of the methods, depending on their emphasis relative to the criterion. Thus, in the comparison of the discussion method with the lecture method of teaching, it is conceivable that the use of a standardized test will be more oriented toward facts and memory work and, thus, favor the lecture method. On the other hand, it is frequently difficult to establish the fairness—that is, the validity—of the test on the basis of which the results of the experiment are to be appraised. It is essential first to define the objectives of the experiment so that the validity of the criterion relative to these objectives can be determined. In the selection of the test to be used, it is also necessary to ensure its adequacy from the standpoint of the purpose of the study.

3. *The Ex-Post-Facto Experiment.* Particularly questionable from the standpoint of control is the *ex-post-facto* experiment,[5] which seeks to identify the antecedents of the differ-

[4] Bertrand Russell, *Philosophy* (New York: W. W. Norton, 1927), pp. 29–30.

[5] Stuart F. Chapin, *Experimental Designs in Sociological Research* (New York: Harper, 1955).

ences noted in existing groups. This is experimentation in reverse: instead of taking groups that are equivalent and exposing them to different treatment with a view to promoting differences to be measured, the *ex-post-facto* experiment begins with existing groups that are different and attempts to trace the antecedents of these differences. The obvious weakness of such an experiment is that it lacks control over past circumstances and cannot isolate the many conditions that may have been involved. For example, if we note that our present civic and business leaders were Boy Scouts proportionately more frequently than non-leaders, we need to realize that their present leadership status is related to a number of factors in addition to membership or non-membership in Scouting. We cannot assume that their previous membership, or non-membership, in the Boy Scouts was a matter of chance, since factors, such as living far from other children, might be involved one way or the other both in their participation in Scouting and in their present community adjustment. Similarly, statistics on the differential earning power of high-school graduates and drop-outs cannot be attributed to a high-school education alone, since a multiple of other factors, such as intelligence, drive, socio-economic background, and many others, may also have played an important part in the present earnings of both the graduate and the drop-out. It would be necessary to equate these factors, perhaps by pairing, before meaningful results could be obtained.

By their very nature, *ex-post-facto* experiments can provide support for any number of different, and perhaps contradictory, hypotheses. If we were to find that unemployed people read more than employed people, we might consider this evidence to support the hypothesis that leisure promotes cultural interests. But, on the other hand, if we were to find they read less, we could suggest as hypothesis that a lack of cultural interest is conducive to unemployment. Such experiments lend themselves to such a large number of crude hypotheses, and are so completely flexible, that it seems to be largely a matter of finding support for hypotheses one wants to hold in the first place. The point is that these hypotheses are not tested, since one cannot test hypotheses on the same data from which they were derived; the evidence simply illustrates the hypothesis. *Ex-post-facto* experiments are, therefore, better considered as sur-

veys useful in the derivation of hypotheses to be tested through more conventional experimental approaches.

Randomization

Our first goal, then, is to control all relevant factors operating in the situation. Since complete control is impossible, however, the investigator must attempt to neutralize the effect of whatever factors have not been adequately controlled by assigning the subjects at random to the various groups under comparison. For example, one might take successive pairs of students, which are equal as far as possible, and assign one member of each pair at random to each of the two groups so that directional differences caused by uncontrolled factors will tend to cancel one another. As a result, the only errors will be those of a random sampling nature, the magnitude of which can be appraised on the basis of the theory of probability.

The effects of a failure to randomize the influence of unequated factors are readily seen in the story of the captain who tested the effectiveness of seasickness tablets by conveniently giving the tablets to his own crew and using the passengers as control.[6] Obviously, the tablets were "effective"; a number of the passengers were seasick—but none of the crew. Similar biases due to a failure to randomize the assignment of subjects to the experimental treatments can be found, for instance, in studies of the effectiveness of a remedial-reading program when an attempt is made to select for the program those pupils who are most likely to profit from it, while the remainder act as a control group. This is apparently what occurred in the Lanarkshire investigation of the value of milk to the health of school children;[7] the teachers, having been allowed some discretion in choosing the experimental group, actually chose children who were in the greatest need of milk. Similar violations of the principles of control and randomization are to be noted in studies in which the experimental groups consist of volunteers or, for that matter, in studies comparing the performance of an experimental group with that of the group on which the norms of a test were derived. In both instances, since the sub-

[6] Wilson, *op. cit.*

[7] Gerald Leighton and Peter L. McKinley, *Milk Consumption and the Growth of School Children* (London: H. M. Stationery Office, 1930).

jects cannot be assumed to have been randomized in their assignment to the groups being compared, a bias in the results is almost sure to exist.

Replication

No matter how carefully one attempts to control all the factors that might influence the results on the basis of which the operation of the independent variable is to be appraised, nor how randomly the methods and the subjects are assigned to the experimental and control groups, slight discrepancies invariably remain. These are taken care of through the replication of the study, which, in essence, is a matter of conducting a number of sub-experiments within the framework of an overall experimental design. Thus rather than comparing a single control case with a single experimental case, the investigator makes a multiple comparison of a number of cases of the control group and a number of cases of the experimental group, all within the same experiment. He might assign equivalent cases to each of the control and the experimental groups at random, and consider the comparison of each pair as an "experiment" in itself. Thus in an experiment involving fifty cases in each of the experimental and control groups, he is really conducting fifty parallel "experiments" in one. In a more elaborate experiment, a number of control and experimental groups, each consisting of essentially equivalent individuals assigned at random to one or another of the two groups are combined within the framework of a single experiment. This would be necessary in order to replicate all aspects of the situation which are not replicated when only two groups are compared—such as the teacher variable in a comparison of two teaching methods.

The precision of an experiment involves a balance between *control, randomization,* and *replication.* Randomness is, of course, essential. Without it, directional differences are likely to occur, the magnitude and direction of which are beyond interpretation. Assuming randomness, precision becomes a function of the degree of control and the extent of replication. Specifically, the precision of an experiment can be increased either by increasing the number of cases in the comparison groups, or by increasing the homogeneity of the samples through a greater degree of control, thus minimizing the influence of the

many variables to which the differences in the outcome might be attributed. This implies that the greater the degree of control, the smaller the sample needed for a given level of precision. In practice, it is therefore a matter of balancing the degree of control that should be exercised against the possibility of relying on a larger number of cases of a somewhat less homogeneous nature. In the Lanarkshire experiment,[8] for example, it was claimed that greater precision in the overall results would have been obtained if the experimenters had used fifty pairs of identical twins instead of twenty thousand cases selected at random.

THE STEPS OF THE EXPERIMENTAL METHOD

The steps of the experimental method are essentially those of the scientific method. For the sake of clarification, they may be listed as follows:

1. *Selecting and delimiting the problem.* The problems amenable to experimentation generally can, and should, be converted into a hypothesis that can be verified or refuted by the experimental data. The variables to be investigated should be defined in operational terms—for example, the scores on a test of acceptable validity.
2. *Reviewing the literature.*
3. *Drawing up the experimental design.* While it should also include a clarification of such basic aspects of the design as the place and the duration of the experiment, this section should place primary emphasis on the questions of control, randomization, and replication. Because of the complexity of an experiment, it is generally advisable to conduct a pilot study in order to ensure the adequacy of the design.
4. *Defining the population.* It is necessary to define the population precisely so that there can be no question about the population to which the conclusions are to apply. College sophomores as experimental subjects, for example, constitute a sample of a population that is, with respect to certain problems at least, extremely ill-defined.
5. *Carrying out the study.* It is necessary here to insist on close adherence to plans, especially as they relate to the factors of control, randomization, and replication. The duration of the experiment should be such that the variable under investi-

[8] W. S. Gossett, "The Lanarkshire Experiment," *Biometrika,* 23 (1931): 398–406.

gation is given sufficient time to promote changes that can be measured and to nullify the influence of such extraneous factors as novelty.

6. *Measuring the outcomes.* Careful consideration must be given to the selection of the criterion on the basis of which the results are to be measured, for the fate of the experiment depends in no small measure on the fairness of the criterion used.
7. *Analyzing and interpreting the outcomes.* The investigator is concerned with the operation of the factor under study. He must be especially sensitive to the possibility that the results of his study arose through the operation of uncontrolled extraneous factors. He must further exclude, at a given probability level, the possibility that his experimental findings are simply the result of chance. In no other area of educational research is the need for competence in statistical procedures so clearly indicated as in the analysis of experimental data as the basis for their valid interpretation.

   Of course, statistics cannot correct faults in the design or overcome inadequacies in the basic data. The investigator must recognize that statistical tools do not relieve the scientist of his responsibility for planning the study, for controlling extraneous factors, and for obtaining valid and precise measurements. Nor does a statistical legerdemain endow with significance a problem which is inherently trivial. It can also be argued that there is limited justification for high-powered statistical refinement in the early exploration of a problem area or in instances where the data involved are essentially crude and imprecise.
8. *Drawing up the conclusions.* The conclusions of the study must be restricted to the population actually investigated, and care must be taken not to overgeneralize the results. The results also pertain only to the conditions under which they were derived, and, since control may have distorted the natural situation, care must be taken to restrict the conclusions to the conditions actually present in the experiment. The investigator must not forget that his conclusions are based on the concept of probability, but, especially, he must not fail to recognize the limitations underlying his conclusions and/or the special conditions that restrict their applicability.
9. *Reporting the results.* The study must be reported in sufficient detail so that the reader can make a judgment as to its adequacy.

EXPERIMENTAL DESIGNS

Experimental designs vary in complexity and adequacy, depending on such factors as the nature of the problem under investigation, the nature of the data, the facilities for carrying out the study, and, especially, the research sophistication and competence of the investigator. Although there are a number of combinations of the various experimental procedures, the basic designs are:

1. the single-group design,
2. the parallel-group design,
3. the rotation-group design,
4. the factorial designs.

These resemble one another from the standpoint of purpose and of their adherence to the principles of scientific experimentation. They differ in the particular manner in which they attack the problem, in the degree of accuracy with which they meet the criteria of control, randomization, and replication, and, of course, in the adequacy of the answers which they are capable of providing.

The Single-Group Design

The single-group experiment is the most elementary and least rigorous design. It consists of comparing the growth of a single group under two different sets of conditions—that is, of subjecting the group successively to an experimental and to a control factor for equivalent periods of time and then comparing the outcomes. The procedure might be listed as follows:

1. Test the group; introduce Method A; test the group again; and note the gains.
2. Allow for a period of transition.
3. Test the group again; introduce Method B; test the group once more; note the gain.
4. Compare the gains in 1 and 3.

This experimental design has a number of limitations that need to be clearly recognized. On the favorable side, it permits an experiment to be conducted by a teacher in his own classroom without assistance, and, on the surface, since the same

group and the same teacher are involved, it seems to make a fair attempt at equating the factors of the ability and background of the subjects and the general characteristics of the experimental situation. On the other hand, this does not necessarily establish experimental control; the students may not be equally motivated by the two methods; nor is the teacher necessarily equally effective and enthusiastic about both The novelty factor is also uncontrolled.

Among the other limitations of the single-group design in an experiment—on teaching methods, for example—are:

1. It assumes that the scale along which growth is to be measured is parallel to the growth curves of the experimental subjects. Since for the short durations, which might be involved in an experiment of this type, we might expect a linear growth, we assume that the gain in raw score of 25 to 45 achieved through Method A is the same as the gain with Method B of 50 to 70. This implies that the learning curve of that period is essentially linear, and that the material presented through Method B is of the same difficulty relative to the present degree of readiness of the students as was the material presented under Method A. It also assumes that the performance of the subjects is in no way affected by a *ceiling* or a *floor* imposed by the instruments used or by the phenomenon of *regressoin toward the mean.*[9] This is frequently difficult to accept.
2. It assumes that, except for differences in the factors being compared, the gains from pre-test to post-test would be the same under both conditions. It assumes, for example, that the

[9] *Regression toward the mean* refers to the tendency for anyone whose performance is extremely low or extremely high on the first test to "regress" toward the mean on the second testing. This generally would favor the first method and penalize the second. A *floor* in tests and measurements refers to the fact that a student's score may be as low as the test allows it to be, but not necessarily as low as his true score—generally because the test is essentially too hard for him. Such a student could make even considerable gain in achievement without it showing on the re-test. Conversely, a student who hit the *ceiling* on the pretest—that is, who scored essentially as high as the test will allow him to score, but not necessarily as high as he could score on a test more in line with his abilities, would have no way of registering any gain during the experimental period. As a rough rule, none of the scores should be close to the zero point (or the score possible through guessing), and none of the scores should be close to the maximum possible —that is, both the pre-test and post-test distributions should be "free-floating" in the sense of being independent of floors and ceilings. Since most abilities are normally distributed, a non-symmetrical distribution of performance should be checked in this connection; a truncated distribution, for example, would be particularly suspect.

practice effects are the same in both cases. This is again difficult to accept since the gains due to practice effects are generally greater from the first to the second testing than from the second to the third.

3. It assumes no undue carry-over in attitude, skill, and information from the first method to the second method. In comparing the teacher-directed and pupil-directed methods of teaching spelling, for example, care must be exercised that the skills learned in the first method are not so fundamental that they vitiate the second.
4. It assumes that the test that is used as criterion is equally valid to the two methods.

The one-group method of experimentation is relatively inadequate, except for purposes of crude experimentation, inasmuch as it fails to comply adequately with the requirements of control and replication. One might go as far as to say that the single-group experiment is not research at all, for it is doubtful if one group's performance on a task can act as its own control for previous achievement under different conditions. Equally faulty is the common one-group experimental design based on the comparison of the growth of the experimental group with that of the norm group on which the test was standardized. This design also would lack control since the norms of the test were not derived under experimental conditions. An extension of the one-group technique consists of having a group alternate from Method A to Method B, back to Method A, and perhaps back again to Method B, at periodic intervals. Such an approach, while undoubtedly a little more dependable, is still subject to essentially the same criticism as attributed to the method in general.

The Parallel- or Equivalent-Group Design

A more adequate experimental design is the parallel- or equivalent-group technique in which the relative effects of two treatments are compared on the basis of two or more groups, which are equated in all relevant aspects. This is essentially the implementation of Mill's canon of difference. In an educational experiment, the groups being compared generally are equated on chronological age, IQ, motivation, sex, general scholarship, general background, and any other factor

considered relevant to the problem under investigation. The basic design of parallel-group experimentation might be represented as follows:

	Experimental	*Control*
1.	Pre-test	Pre-test
2.	Experimental factor	Control factor
3.	Final test	Final test
4.	Comparison of gains	

Where equivalence of the groups has been established, such a design can generally provide reasonably dependable conclusions relative to the operation of a given factor, and, of course, the greater the control exercised, the greater the precision of the results. Theoretically, the equivalence of the two groups is best established through the matched-pair technique, which consists of pairing individuals on relevant factors and assigning a member of each pair to the experimental and control groups at random. A further refinement would be to use identical twins who, because of their similarity, would generally provide much greater control than would an equal number of individuals selected at random or matched on the basis of general equivalence.

On the other hand, an equivalent-group design based on matched pairs suffers from obvious practical difficulties. Despite the fact that there tends to be a correlation among the usual bases of pairing—for example, IQ, mental age, scholarship, and pretest scores—invariably only a fraction of the members of a population can be paired on a multiple basis with any degree of precision. In a school situation where it is possible to shift students from one class to another, a few more pairs can generally be located, but invariably a substantial segment of each class matches no one in the other group, and the investigator is almost forced to exclude them from the study. This not only reduces sample size but also may introduce artificiality into the situation by reducing the class size below normal enrollment. If the unmatched students are simply allowed to remain in the class but are not included in the experiment, they introduce a disturbing effect which can invalidate the experiment. In an extended study there is also a possibility that subjects will drop out from one or the other of the two groups,

forcing the removal of their mates—thus reducing sample size and decreasing the precision of the experiment.

From an administrative point of view, it is more convenient to match groups rather than pairs where two groups are considered equivalent when they have equal means and equal standard deviations in each of the variables considered relevant to the purposes of the investigation. Sometimes this is done simply on the basis of chance; it is assumed that if one takes a sufficiently large number of cases at random—or a number of classes to which students have been assigned at random—factors that may be involved in the experiment will simply equate themselves. This is risky, particularly in dealing with small groups. The equivalence of the groups should be tested, and adjustments made where indicated.

The *matched-group design,* where its conditions are fulfilled, is relatively adequate. It has certain administrative advantages over matched pairs in that it permits the full use of the total groups, even when the size of the two groups is not equal. On the other hand, it does not have the same degree of precision as the matched-pair design. Furthermore, it is not always possible to find two or three prearranged groups (classes, for instance) to be equivalent in a number of respects. To make matters worse, even if such an ideal situation could be attained —or arranged through a reshuffling of the classes—it is doubtful that it would remain that way for an experiment of extended duration. Drop-outs, which are likely to occur, will disturb the equivalence of the groups and necessitate some readjustment in the equating.

Analysis of Covariance. The more sophisticated and adequate method of handling the situation is to rely on statistical equation of the groups through *analysis of covariance.* This technique is a procedure which permits statistical adjustments to be made in the dependent variable in order to compensate for any lack of equivalence between the groups in the independent variables. For example, in the study of academic growth associated with different teaching methods, analysis of covariance would permit an adjustment to be made in the test gains for slight differences that might exist in the IQ levels of the groups. Actually, the procedure is simply an extension of the concepts of regression and correlation. It is not beyond

the capabilities of the average graduate student to apply, if not to understand from the standpoint of mathematical derivation. When adjustment needs to be made for only one of the independent variables, even the labor involved is relatively minor.

In practice, the investigator first attempts to obtain general equivalence of the groups in such major factors as IQ, since, obviously, no statistical technique will permit a precise adjustment for the effects of IQ on academic achievement when the two groups consist one of morons, and one of geniuses. If only a small discrepancy is involved—due either to failure to equate the groups completely in the first place or to drop-outs—an adjustment can be made through analysis of covariance that will permit the comparison of the two groups on the dependent variable, as if they had been equated.

The Rotation-Group Design

When the experimental and control groups are only approximately equivalent in relevant factors, it may be possible to conduct the investigation by rotating the groups at periodic intervals. For example, Groups A and B might use Methods X and Y, respectively, for the first half of the experiment and then exchange methods for the second half. A comparison would then be made of the relative gains of each of the groups under the two methods. This approach is essentially an extension of the one-group design, but it minimizes some of its weaknesses and permits a somewhat more rigorous interpretation of the results. For example, if Method X proves to be superior when used by both Group A and Group B, the answer is fairly clear. If, on the other hand, Method X should prove superior when assigned to Group A but inferior when assigned to Group B, it might be suspected that Group A is noticeably superior to Group B in its ability to achieve regardless of method, or that some factor was not controlled—for instance, transfer from one method to the other, the enthusiasm of the teacher, and so on. A more adequate design, which would incorporate the advantages of both the equivalent-group and the rotation-group design, is the rotation of equivalent groups. A further extension might have equivalent groups rotated back and forth from one method to the other a number of times.

Synthesis. All three of these designs have potentiality. Furthermore, they can be extended to include any number of groups and, of course, any number of levels of each of the variables being considered. Macomber and Siegal,[10] for example, used the parallel-group method to compare the relative performance of students in large TV classes, in large auditorium classes, and in small regular classes. Their particular worth would have to be evaluated in the individual case in the light of the criteria set forth in this chapter. It is particularly important in a study of teaching methods, for instance, to have a number of classes of each of the experimental and control groups in order to fulfill the conditions of replication with respect to the teacher variable. Otherwise, the study is simply a matter of comparing one teacher's effectiveness with that of another. And, of course, using the same teacher in the two groups does not solve the problems of replication and control, since he may not be equally competent in the two situations.

Unfortunately, all these designs are based on the law of the single variable and, as such, are of limited value in promoting education as a science, except in problems of relatively limited complexity. This is not to minimize their worth. Undoubtedly, a number of problems can be investigated through such procedures. On the other hand, the fact that they tend to promote artificiality in complex situations cannot be overlooked, and their use, while not outdated, is automatically restricted to specific problems for which they are appropriate. In fact, many of the inconclusive results and conflicting and contradictory outcomes of experimentation recorded in journals are in no small measure related to the inappropriateness of the monistic experimental design in a situation where a more complex design is required.

Factorial Designs

At a more adequate level, particularly from the standpoint of the complex problems with which the social sciences are concerned, are the *factorial designs*—or *multivariate analy-*

[10] Freeman G. Macomber and Laurence Siegal, "Study in Large-Group Teaching Procedures," *Educational Record,* 38 (July 1957) : 220–9.

sis—which permit the simultaneous evaluation of the effects of a number of factors taken singly and in interaction with one another. These modern experimental designs have made educational problems more and more amenable to investigation in their sub-aspects, as well as in their entirety. It is possible, for instance, to investigate teacher effectiveness through a factorial design incorporating such factors as experience, degree status, scholarship, personality, and so on.

Among the more common approaches to multivariate analysis are Latin squares, Greek squares, randomized blocks, and confounding designs. These procedures are obviously too technical for treatment here. Yet in view of the crucial role of multivariate analysis in the solution of educational problems, they warrant the careful attention of both consumers and producers of research. The students interested in the mechanics of computation can find many examples reported in more or less complete detail in doctoral dissertations and in textbooks in advanced statistics. A particularly thorough treatment of a 2 x 3 x 3 x 3 arrangement dealing with achievement in high school biology is given by Johnson and Tsao.[11] The example incorporates analysis of covariance introduced as a means of controlling the factors of mental age and pre-test scores and, thus, increasing the sensitivity and precision of the test of the various hypotheses. The same authors[12] present a 4 x 7 x 2 x 2 x 2 design dealing with the ability of individuals to sense differences in the weight of objects.

A design commonly used in biostatistical experiments is the *Latin square,* which might be conceived as a form of rotation design in which a variety of treatments are assigned to different groups under different experimental conditions. In a 4 x 4 Latin square, for example, four different teaching methods may be randomized among four different motivational approaches so that every combination of method and motivation is represented by at least one group. Any number of these patterns can be devised to incorporate any number of variables.

A particularly interesting experimental design is the

[11] Palmer O. Johnson, and Fei Tsao, "Factorial Design and Covariance in the Study of Individual Educational Development," *Psychometrika,* 10 (June 1945) : 133–62.

[12] Palmer O. Johnson, and Fei Tsao, "Factorial Design in the Determination of Differential Limen Values," *Psychometrika,* 9 (June, 1944) : 107–44.

Johnson-Neyman technique[13] which permits not only the determination of the relative effectiveness of variables, but also the region for which a given variable is effective. For example, rather than compare the relative effects of gifted teachers and less gifted teachers on student growth, it might be possible to derive the conclusion that gifted teachers are significantly more effective in promoting the academic growth of gifted children, and that less gifted teachers are significantly more effective in promoting the growth of duller children than any other combination of teacher and pupil ability.

Although multivariate analysis is not a panacea for unlocking the secrets of the many problems with which education must cope, it constitutes, in most cases, a more effective approach than the traditional method currently in common use. On the other hand, it tends to be unwieldy. For example, an adequate comparison of five factors at each of two levels, would require a minimum of thirty-two different observations—that is, five primary effects, ten first-order interactions, ten second-order interactions, and five third-order interactions, and one fourth-order interaction involving all five factors.[14] Because of their complexity such experimental designs require knowledge of statistics beyond the average teacher's competence. This is unfortunate in view of their importance to the advancement of education. The solution to the dilemma is probably a matter of placing the responsibility for the development of education as a science in the hands of the professional research worker with a good background in research and statistical methods. It would also mean that each school system has a definite obligation to provide this leadership.

Causal-Comparative Studies

The problems of education, as we have seen, are frequently complex. Whether we consider teacher-effectiveness, under-achievement, or delinquency, we find that they incorporate a multiplicity of causal factors, contributing factors, and precipitating factors, as well as an unlimited number of other

[13] Palmer O. Johnson, and Leo C. Fay, "The Johnson-Neyman Technique, its Theory and Application," *Psychometrika*, 15 (December 1950) : 349–67.

[14] It should be noted that third and fourth order interactions often are relatively meaningless from the standpoint of grasping them intellectually or doing anything about them practically.

elements of varying degrees of relevance—all operating in different degrees of interaction. Obviously, any attempt to investigate the effect of any one of these aspects through a series of simple experiments—in isolation of the other agents in the situation—is not likely to yield anything but a series of half-answers, which is essentially the same as no answers at all. On the other hand, the use of multivariate analysis is limited considerably by the unwieldiness of the procedure.

The predicament stems in part from the relative lack of clarity with which many of the so-called complex problems situations with which education must cope have been defined. In our present development of educational science, we do not have a sufficient understanding of many of the more complex educational problems as they exist in actuality. Until their various aspects have been structured more definitively, and the number of their relevant factors is brought down to manageable size, their investigation through factorial designs is relatively impractical and, in some instances, impossible.

A common approach to structuring the field in order to gain greater insight into complex situations is to select two groups at opposite ends of the continuum in order to identify the factors on the basis of which one group can be distinguished from the other. Research into the contrasting characteristics of juvenile delinquents and non-delinquents, for example, has shown the former to be more independent, extrovertive, vivacious, impulsive, aggressive, adventuresome, and, of course, more lacking in self-control.[15]

This approach, sometimes known as *causal-comparative,* obviously is difficult to use effectively. It places a particularly great burden on the imagination and insight of the investigator to identify the crucial aspects of the situation. First, he obviously cannot consider everything; nor would he want to, since it is essential to keep the number of variables to be analyzed to a minimum, if confusion is not to result. Yet to the extent that he leaves out factors that are relevant, his solutions will be lacking. This is complicated further by the fact that crucial factors are frequently subtle. Good teachers, for example, differ from poor teachers principally in contributing

[15] Sheldon Glueck, "The Home, the School, and Delinquency," *Harvard Educational Review,* 23 (1953) : 17–32.

rather than in critical factors. An interesting possibility in this connection is the use of the electronic computer in permitting the analysis of large masses of data that might be gathered in a study of this kind. A somewhat more scientifically oriented approach would be to rely on factor analysis to reduce the number of variables to be considered in a study by identifying the fundamental psychological dimensions that underlie the operation of more superficial traits.

LABORATORY EXPERIMENTATION

Modern statistical advances have refuted the belief that only laboratory experimentation can provide the control necessary to obtain precise results, for, as we have seen, multivariate analysis permits the rigorous investigation of the multiple aspects of a complex problem in its natural setting. Nevertheless, though modern factorial designs have lessened the need for laboratory experimentation, there are still many instances in which, because of the need for more intensive investigation of a small segment of an overall problem situation, laboratory experimentation seems indicated. The contributions of laboratory experimentation in providing both insight into a complex problem and hypotheses to be tested under more normal conditions can be seen in the early work of Thorndike[16] in the psychology of learning.

Because they permit greater control over the operation of extraneous factors, laboratory studies usually are more precise than the corresponding field investigations. On the other hand, the laboratory situation is automatically more artificial, and overemphasis on control may not only produce meaningless results, but also may focus the attention of investigators on trivial problems and sacrifice the investigation of more significant variables. This would, of course, be a dubious bargain for the accuracy which laboratory experimentation provides.

Laboratory experimentation depends greatly on instrumentation for increasing the precision of observation, and, of course, with modern technological and psychological advances, great strides have been made in this area. We must not, however, overlook the fact that the crucial determinant of the precision and accuracy of any experiment is the investigator,

[16] See Chapter 15.

rather than the extensiveness of his equipment, and it does not necessarily follow that the more extensive the equipment, the more adequate the conclusions that are derived.

The laboratory method is, of course, most appropriate to the physical sciences, where the conditions underlying the law of the single variable are more likely to be fulfilled. Since laboratory experimentation is a precise technique, its use is generally restricted to problems of a specific and limited nature. The laboratory situation is made deliberately artificial with respect to the precision with which the non-experimental factors can be controlled in order to derive more precisely the relationships of the phenomena under investigation. Undoubtedly, it can, at times, provide a better picture of the operation of a given variable as it functions by itself than can less controlled investigations. Once the operation of a variable in isolation is known, it may be easier to understand more adequately its operation in interaction with the other aspects of the overall situation.

On the other hand, because of the artificiality of the laboratory setting, any conclusion based on laboratory experimentation—either in the testing of drugs or in the learning of animals—has to be verified in the field before being implemented or extended to the general population. The investigator, therefore, must never overlook the exploratory nature of laboratory experimentation and the consequent restriction of any conclusion he reaches to the role of hypotheses. This is not to deny the crucial role played by laboratory experimentation both in the exploratory stages of locating something that is likely to work and in the final stages of deriving precise laws.

EVALUATION OF EXPERIMENTATION

Although experimentation has been largely responsible for the tremendous development of the physical and biological sciences, experimentation in the field of education has not fulfilled its earlier promise. Not only has it failed to provide little more than superficial answers to many of the problems with which educators are faced, but it has provided conflicting and contradictory outcomes. As a result, many educators have come to feel that experimentation is not suitable for the investigation of many important educational problems, and that the an-

swers to their problems will have to come from judgment, experience, common sense, and consensus. This does not imply that no progress has been made. A number of important studies have been conducted since Rice's original investigation,[17] and many of them have had a great influence on educational practice.

That early progress was relatively slow can be understood in light of the difficulties—lack of statistical tools, of psychological tests, and of research know-how—faced by early investigators. These limitations have been alleviated by modern developments in the statistical and psychological areas, though, unfortunately many of the newer advances are apparently beyond the present competence of the bulk of the profession. It also needs to be recognized that there are inherent in educational research certain limitations that make for immediate difficulties. For instance, we are still lacking basic tools for the appraisal of such important educational factors as attitudes and motivation. Furthermore, in contrast to the physical sciences where the results are relatively immediate, outcomes in the field of educational research frequently emerge only slowly and are affected by so many variables that it is difficult to assign a certain outcome to a given antecedent. It is also true that gains with respect to one criterion frequently are offset by losses with respect to another. The very fact that we deal with human subjects poses special problems of a more complicated nature than those confronting the physical scientist. Not only are human subjects not as easily manipulated and controlled as are non-human subjects, but human beings also tend to react to the experiment—with increased motivation, for example, —and thus, vitiate the results.

As we have seen, the basic concept of control immediately creates complications in a field whose variables are invariably complex, and, in general, the more precise an experiment is, the more artificial it is, and the more meaningless are its results. Much of the lack of progress in educational experimentation can be attributed to our futile attempts to make our experiments conform to the law of the single variable, and to our failure to see the inappropriateness of the univariate approach for dealing with the complex problems facing educa-

[17] See Chapter 15.

tion. The alternative, multivariate analysis, on the other hand, is relatively complicated.

Thus far, too little experimentation has been done in education. Educators as a group tend to shy away from experimentation, especially of the long-range complex variety needed to provide adequate answers to educational problems. Inasmuch as the teachers who are facing the problems rarely have the training necessary to solve them, there is a tendency for the experiments, which are carried out, to lack continuity. Even teachers and administrators with advanced degrees are rarely sufficiently well-versed in research and statistical techniques to conduct such research. University faculties in education, in contrast to those in the fields of psychology and the physical and biological sciences, are relatively unoriented toward experimentation. Catalogs of colleges of education suggest that the training of educators in experimental and statistical methods is relatively meager; a student interested in such fields would have to take his training in the department of psychology. In other words, in a field in which we need particular competence, if we are to deal with our problems, we are lacking even in basic skills.

Furthermore, the experimentation that is conducted is frequently inadequate, if not incorrect. Norton and Lindquist, in their recent review of educational experimentation,[18] point to numerous and damaging recurring errors in experimental design and in the analysis of experimental results. Not only is there inadequate use of advanced statistical procedures, despite the fact that multivariate analysis has been in existence for thirty years, but the simple studies that have been conducted often fail to comply with elementary conditions of control, randomization, and replication. Frequently, such relevant factors as teacher effectiveness are overlooked and become confounded with effects of the variable under study. Among the other common errors found in educational experimentation are the invalidation of a criterion through the experimental design, failure to consider the assumptions underlying the procedures used, inadequate and non-representative sampling,

[18] Dee W. Norton, and Everett F. Lindquist, "Applications of Experimental Designs and Analyses," *Review of Educational Research*, 21 (December, 1951) : 350–67.

and carelessness in defining the population, and delimiting of the generalizations derived to the population involved. In fact, as pointed out by Stanley, the professional literature is virtually devoid of any well-controlled experimental studies conducted in the classroom.[19] This situation needs to be remedied: the tools and the procedures are available and the problems are there; it is simply a matter of orienting ourselves to implementing the use of the one to the solution of the other.

CASE STUDIES

Closely related to experimentation is the *case study* whose purpose also is to identify the antecedents responsible, in a direct or indirect causative way, for the occurrence of such phenomena as reading disability, maladjustment, immaturity, and delinquency. Actually, the case study resembles almost all other types of research. It borders on historical research, for instance, in the sense that the present case can be understood only in view of its past. It is closely related to documentary research in that it deals with living individuals in their present social environment. Case studies resemble survey studies in that they are concerned with the present status of phenomena. They differ from survey studies, however, in that the determination of status is only a secondary aspect in the situation; the more fundamental question is discovering how it got that way.

Case studies, as the term is generally used, differ from experimentation in that they display a greater element of subjectivity and intuition and, as they are usually conducted—that is, in a guidance rather than a research setting—are generally oriented toward the solution of a particular problem at the individual level, rather than toward the derivation of generalizations that have scientific validity. Although case studies constitute the most comprehensive means of studying the whole child, a distinction needs to be made between their guidance and their research functions. Undoubtedly, case studies used for guidance purposes can lead to the derivation of relationships that have a bearing on research and vice versa. Yet, in the strict sense of the term, research is concerned with the deriva-

[19] Julian C. Stanley, "Controlled Experimentation in the Classroom," *Journal of Experimental Education,* 25 (March 1957) : 195–201.

tion of generalizations that apply beyond the individual case, and case studies become research only when they are able to supply such generalizations. Consequently, the case study of Johnny, undertaken for the purpose of helping him adjust to the school situation, has limited bearing on research.

As a specialized research technique, the case study presents certain difficulties. The study of five or six cases, even of individuals displaying a high level of homogeneity, is not likely to provide any but the most tentative and crude generalizations. Such studies can, of course, provide insights into the dynamics of human behavior and its antecedents, but, unless a sufficient number of cases is taken to permit the isolation of crucial factors, the extent to which case studies can lead to valid generalization is extremely limited. In general, the case study might be considered primarily a clinical procedure, and only secondarily a research technique. It probably makes its greatest contribution to the advancement of science as a source of hypotheses to be verified by more rigorous investigation.

Case-Study Data

The major problem of the case study is essentially the same as that of the historical method—that is, obtaining dependable data from which valid interpretations are to be derived. Not only are gaps bound to exist in the data, but the data that are available generally have not been collected for the purpose of elucidating the present problem, and invariably they are incomplete, inaccurate, and otherwise inadequate.

The investigator's first task is to gather data that will supply a relatively complete picture of the case. Generally, this involves the use of observation, interviews, tests, and other data-gathering devices and techniques designed to provide information on the individual's life history, his health history, his scholastic history, his home and community background, and any other aspect of the situation that might clarify the present problem. This information will have to be checked for accuracy; much of it will be relatively unverifiable except on the basis of general plausibility.

Along with the gathering and verifying of data, the investigator must devote himself to the even more important and demanding task of interpreting the data in the light of the

present predicament of the individual, with a view to a diagnosis of the case and a prognosis of its likely disposition. This calls for an insight into the past and present situation which is sufficient to permit the synthesis of the relevant aspects of the data with the present problem. This is frequently difficult to achieve since most situations covering a substantial portion of the individual's life are too comprehensive to permit a complete investigation of every aspect.

The problem is complicated further by the fact that these data cover a multitude of different facets of the individual's background, and they generally do not lend themselves to easy statistical synthesis. This is not to glorify the quantitative approach, for the more important problems in education are not necessarily those that lend themselves best to quantification. The implication is only that the case-study method thus far has relied too heavily on the investigator's judgment, if not intuition.

In order to make sense out of the mountains of data which he may accumulate, the investigator must make use of hypotheses derived from the superimposition of his general theoretical orientation on relatively incomplete data covering a complex situation. This calls for insight into human nature as it exists in its sociological context, and it always involves obvious risks of error. The investigator's mind-set may blind him to certain significant aspects of the situation, and his general orientation determines to a large extent the relevance he attributes to the data he collects and the interpretation he gives them. As a result he may build out of his personal experience and perspective a case which has relatively little foundation in actuality.

Rationale of Case Studies

Case studies generally involve the co-operation of a number of investigators pooling their resources toward the diagnosis, the prognosis, and perhaps the treatment of a problem. In the guidance of a child who is displaying anti-social behavior, for example, a team, consisting of the school psychologist, teachers, guidance workers, social workers, and other interested persons, pools its information and insights in order to gain an understanding of the case. Eventually a diagnosis is reached

and remedial steps are prescribed. The latter validates the diagnosis; if the treatment alleviates the symptoms, it can be assumed that the source of difficulty has been properly identified, and that the problem is probably on its way toward disappearance. Conversely, if the symptoms persist, it might be suspected either that the cause of the difficulty has not been properly identified or that an improper inference has been made about the treatment implied by the diagnosis.

The case study is, of course, a fundamental technique in medicine, where diagnosis and treatment as outlined above are standard procedures. There is, however, need to make a distinction between the problem as it exists in medicine and as it exists in education and psychology, where diagnosis is frequently more complex and more risky. In the field of medicine, the diagnosis is relatively clear from the symptoms: a slight fever, a swelling of the lower jaw, and so on, spells mumps with relative certainty. Once the identification is made, the treatment is generally prescribed, and the cure follows in rather short order so that the diagnosis can be validated. In education and psychology, on the other hand, the problem is not so simple. First, the symptoms rarely identify the cause except in a very tentative way. The child who is anti-social may be anti-social for any one of many reasons, ranging from feelings of rejection to feelings of hunger. Consequently it is frequently necessary to collect a great deal of information about the individual and to pool the insight of a number of experts to arrive at a sound diagnosis.

In the social sciences, the problems of devising remedial procedures and of implementing the solution are also more difficult. Failure of the home to co-operate, for instance, may preclude a cure. Consequently, when treatment does not work, it is difficult to know who is to blame for the failure—or even to know if a failure is involved. Improvement frequently is slow, and even the most correct technique can aggravate the symptoms while reorganization is taking place, which may cause the person in charge to give up the treatment just as improvement is about to occur. It is also difficult to attribute success to any one cause. In reading, for instance, it is common to attribute the child's improvement in reading to the remedial procedures, when it may stem, in part at least, from the greater

attention the child is receiving. Thus, even when a cure is effected, the investigator may not have learned very much from a scientific point of view.

The Steps of the Case Study

If it is to be accepted as a scientific technique, the case study must follow essentially the same steps and meet essentially the same criteria as do the other research methods. On the other hand, it presents a number of problems which are relatively unique, either in kind or in degree. These are probably best considered in connection with the steps through which such a study must proceed.

1. The first step of the case study is obviously the selection of the cases which exemplify the problem area under consideration. There is especially a need for typical cases—that is, not a random sample of the general population but a random sample of cases considered representative of the problem under investigation. The sample should be large enough to permit the derivation of valid generalizations. This often presents a problem. Since case studies cover many facets of the total picture and extend over a long period of time, and are therefore time-consuming and costly, it is common practice to restrict the study to the thorough investigation of a few cases. This, of course, raises the question of the representativeness of such small samples and of the degree of certainty with which the results can be generalized to the alleged population.

2. The collection of data on the individual cases must be guided by some tentative hypothesis. Some of the data will be readily available from records and will pose no problem of collection. There will, however, be the question of verification and interpretation. Generally when these data were collected, present needs were not anticipated, and, as a result, the data were probably not collected and recorded systematically enough to be dependable and understandable in the context of the present problem. The cumulative record, for example, may include test scores recorded without date or identification. Some data will be incorrect or invalid or will include information that, though correct, is misleading. Some of the data will have to be collected from the community where the emphasis is often on hearsay, on the atypical and, of course, on memory. Some will have to come from parents who may not have insight into whether the child,

was insecure as a baby or whether he had unusual troubles at school, for example.[20]

3. An important step in the case study is the derivation of hypotheses or tentative diagnoses of the likely antecedents of the difficulty. Generally this is followed by some mental elaboration of the diagnosis in the light of the data already available and, where necessary, by the collection of further data.

4. Along with the diagnosis comes the suggested treatment and the prognosis of the likely response of the difficulty to such treatment, judged in the light of the severity of the case and the environmental circumstances under which the cure is to be effected. There is a definite need here for insight into the dynamics of human behavior as they operate in a sociological setting, and effectiveness in case studies generally calls for considerable training in the areas of psychology and sociology. A common practice, when case studies are to be used in guidance, is to implement multiple remedial procedures simultaneously on the assumption that they will do no harm. From a research point of view, such an approach does not provide the generalizations which science requires in order to deal with subsequent cases.[21]

5. The final step is the follow-up of the case from the standpoint of its response to treatment. This constitutes a test of the validity of the diagnosis.

SUMMARY

1. Experimentation is undoubtedly the most scientifically sophisticated research method. It is a refined technique capable of providing precise answers to precise problems. Its use, therefore, is restricted to the later stages of the investigation of a problem, after it has become sufficiently structured—as a result of investigation through more flexible approaches—to permit the derivation of specific hypotheses which can then be submitted to experimental test. The experiment is conducive to both economy and precision because it stages the occurrence of a phenomenon under conditions

[20] There is a problem of ethics involved in collecting data of a fairly personal nature and of varying degrees of accuracy so that the investigator may gain more information, and misinformation, about an individual than the individual has about himself. These data must be kept confidential, especially if there is any possibility of their being used to the detriment of the individual. It is generally unwise, for example, to collect data on such matters as race or religion, unless they are vital to the investigation. On the other hand, as in all research, records should generally be as complete as possible in order to permit possible reanalysis.

[21] In practice, it is frequently inadvisable to postpone treatment indefinitely while a one-at-a-time approach is being tried. Furthermore, certain remedial procedures may not work in isolation.

as free as possible from confounding co-occurrences, and thus permits the more precise and rigorous allocation of the occurrence of the phenomenon to the operation of the experimental factor.

2. In the earlier conception, experimentation was based on the law of the single variable and was oriented toward the discovery of cause-and-effect relationships of the one-to-one variety. This interpretation was not only unnecessarily narrow and incapable of fulfillment, but it also introduced artificiality in the situation and thus tended to vitiate the very purpose of the investigation.

3. If an experiment is to provide dependable answers, it must be self-contained—that is, it must provide the basis for the interpretation of its results. In order to do this, the experiment must comply with three basic and interrelated conditions: control, randomization, and replication.

4. The basic condition underlying experimentation is control, without which there is no way of knowing whether the results noted are due to the operation of the variable under investigation or to some extraneous factor. This calls for having one or more "control" groups to act as a point of reference in evaluating the effects of the experimental factor.

5. Since control of all extraneous factors operating in the situation is impossible, it is necessary to assign the subjects at random to the experimental and the control groups to neutralize the effects of whatever variables have not been adequately controlled. Of course, no matter how carefully extraneous factors are controlled, nor how carefully subjects and treatments are randomized, slight discrepancies between the two groups are still likely to exist because of the operation of chance. It is therefore necessary to replicate the comparison with regard to each of the relevant variables.

6. The steps of the experimental method are essentially those of the scientific method. The design of the experiment is of special concern, particularly from the standpoint of the establishment of control. The adequate analysis of the results of an experiment—especially of the more complex variety—calls for some understanding of advanced statistical procedures.

7. Experimental designs range from the relatively inadequate single-group design to the more sophisticated factorial designs. Despite the obvious limitations of the monistic approach, the parallel-group experiment has probably been the most commonly used experimental design to date.

8. Laboratory experimentation is designed to provide a precise answer to a restricted aspect of a specific problem under rigorously controlled—and relatively artificial—conditions not possible in a field experiment. The resulting insights can then be transferred to the more adequate investigation of the overall phenomenon in the more realistic field setting. The method is more common to the physical sciences and medicine.

9. Unfortunately, educational science has not attained the stage of development at which many of its significant problems are amenable to experimental procedures of the monistic approach which has dominated educational experimentation to date. On the other hand, though much more realistic and adequate for dealing with complex variables, multivariate analysis is relatively complicated and unwieldy.

10. The multiplicity of factors that may be involved in the occurrence of a complex phenomenon automatically makes its investigation through multivariate analysis laborious and complicated. Before multivariate analysis can be used effectively in the solution of the complex problems characteristic of education, there is need for clarification of their nature and the structuring of their components into a somewhat more fundamental organization. The causal-comparative approach, which is designed to identify the contrasting characteristics between representatives of the two extremes of a given phenomenon, can sometimes provide insight into which factors need to be considered and which can perhaps be ignored.

11. The case study is also concerned with the antecedents of such relatively complex phenomena as delinquency or reading disability. The major difficulty in its conduct generally centers around the accumulation of accurate background data and their interpretation in the light of the present predicament of the individual. The technique is most frequently used in a clinical rather than a research setting; it becomes research only to the extent that it permits the derivation of generalizations of relatively broad applicability. In general, case studies serve their greatest research functions through the suggestion of hypotheses that can then be investigated more adequately by more rigorous techniques.

PROJECTS and QUESTIONS

1. What is the present status of some of the classical studies conducted in education? Where have they been repeated? What changes have been suggested in their conclusions?
2. Re-design one of the classical studies referred to above. Identify its strength and its weaknesses and the changes to be made in its improvement.

SELECTED REFERENCES

Experimentation

BARR, ARVIL S., *et al. Educational Research and Appraisal.* New York: Lippincott, 1953.

BORING, EDWIN G. "The Nature and History of Experimental Control," *American Journal of Psychology,* 67 (December 1954): 573–89.

CHAPIN, F. STUART. "Experimental Design in Sociology: Limitations and Abuses," *Social Forces,* 29 (October 1950): 25–32.

———. *Experimental Designs in Sociological Research.* Rev. ed.; New York: Harper, 1955.

COCHRAN, WILLIAM G. and COX, GERTRUDE M. *Experimental Designs.* 2nd ed.: New York: Wiley, 1957.

CONANT, JAMES B. *The Growth of the Experimental Sciences.* Cambridge: Harvard University Press, 1949.

CORNELL, FRANCIS G. and MONROE, WALTER S. "Experiment," in Monroe, Walter S. (ed.), *Encyclopedia of Educational Research.* 2nd ed.; New York: Macmillan, 1950.

EDWARDS, ALLEN L. *Experimental Design in Psychological Research.* rev. ed.; New York: Rinehart, 1960.

EDWARDS, ALLEN L. and CRONBACH, LEE J. "Experimental Design for Research in Psychotherapy," *Journal of Clinical Psychology,* 8 (January 1952): 51–9.

ENGELHART, MAX D. "The Analysis of Variance and Covariance Techniques in Relation to the Conventional Formulas for the Standard Error of a Difference," *Psychometrika,* 6 (August 1941): 221–33.

FISHER, RONALD A. *The Design of Experiments*, 6th ed.; New York: Hafner, 1951.

FOSHAY, ARTHUR W. and HALL, JAMES A. "Experimental Moves into the Classroom," *Teachers College Record,* 51 (March 1950): 353–9.

FUND FOR THE ADVANCEMENT OF EDUCATION. *Decade of Experiment, 1951–61.* New York: The Fund, 1961.

IVERSON, MARVIN A. and REUTER, MARY E. "Ego Involvement as an Experimental Variable," *Psychological Reports,* 2 (1956): 147–81.

JOHNSON, PALMER O. and FAY, LEO C. "The Johnson-Neyman Technique, Its Theory and Application," *Psychometrika,* 15 (December 1950): 349–67.

JOHNSON, PALMER O. and NEYMAN, JERZY. "Tests of Certain Linear Hypotheses and their Application to Some Educational Problems," *Statistical Research Memoirs,* 1 (1936): 57–93.

JOHNSON, PALMER O. and TSAO, FEI. "Factorial Design and Covariance in the Study of Individual Educational Development," *Psychometrika,* 10 (June 1945): 133–62.

KEMPTHORNE, OSCAR. *The Design and Analysis of Experiments.* New York: Wiley, 1952.

———. "The Design and Analysis of Experiments with Reference to Educational Research," in Collier, Raymond O. and Elam, Stanley M. (eds.), *Research Designs and Analysis.* Second Annual Symposium. Bloomington: Phi Delta Kappa, 1961.

LINDQUIST, EVERETT L. *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton-Mifflin, 1953.

MCCANDLESS, BOYD R. and SPIKER, CHARLES C. "Experimental Re-

search in Child Psychology," *Child Development,* 27 (March 1956) : 75–80.

NELSON, WALLACE B. "Experiment with Class Size in the Teaching of Elementary Economics," *Educational Record,* 40 (October 1959) : 330–41.

PATTERSON, CECIL H. "Matching vs. Randomization in Studies of Counseling," *Journal of Counseling Psychology,* 3 (Winter 1956) : 262–72.

RAY, WILLIAM S. *An Introduction to Experimental Design.* New York: Macmillan, 1960.

RYAN, THOMAS A. "Multiple Comparisons in Psychological Research," *Psychological Bulletin,* 56 (January 1959) : 26–47.

SHANNON, JOHN R. "Experiments in Education: A New Pattern and Frequency of Types," *Journal of Educational Research,* 48 (October 1954) : 81–93.

STANLEY, JULIAN C. "Controlled Experimentation in the Classroom," *Journal of Experimental Education,* 25 (March 1957) : 195–201.

TIEDEMAN, DAVID V. "Experimental Method," in Harris, Chester W. (ed.), *Encyclopedia of Educational Research.* 3rd ed.; New York: Macmillan, 1960.

UNDERWOOD, BENTON J. *Psychological Research.* New York: Appleton-Century-Crofts, 1957.

WILSON, E. BRIGHT. *Introduction to Scientific Research.* New York: McGraw-Hill, 1952.

Genetic Research

CORNELL, ETHEL L. and ARMSTRONG, CHARLES M. "Forms of Mental Growth Patterns Revealed by Re-analysis of the Harvard Growth Data," *Child Development,* 26 (September 1955) : 169–204.

DEARBORN, WALTER F. and ROTHNEY, JOHN W. M. *Predicting the Child's Development.* Cambridge: Sci-Art Publishing Co., 1941.

GESELL, ARNOLD S. *Studies in Child Development.* New York: Harper, 1948. (See Chapter 15 for other references.)

KROGMAN, WILTON M. *The Physical Growth of Children: An Appraisal of Studies, 1950–55.* Lafayette: Child Development Publications, 1956.

MERRY, FRIEDA K. and MERRY, RALPH V. *The First Two Decades of Life.* 2nd ed.; New York: Harper, 1958.

NEILON, PATRICIA A. "Shirley's Babies after Fifteen Years; Personality Study," *Pedagogical Seminary,* 73 (December 1948) : 175–86.

PRESSEY, SIDNEY L. and KUHLEN, R. G. *Psychological Development Through the Life Span.* New York: Harper, 1957.

TERMAN, LEWIS M. *et al. Genetic Studies of Genius.* 5 vols. Stanford: Stanford University Press, 1925, 1926, 1930, 1947, and 1959.

Outside of womankind, few topics have intrigued and tormented mankind more than the problem of predicting the future.

NICHOLAS A. FATTU

13 Predictive Methods

Man is forever predicting the likely outcomes of his efforts. In fact, both at the personal and at the professional level, man's behavior implies some degree of expectancy. We invest our money because we expect to make a profit; we enroll in school because we expect to get a degree; we work for a degree because we expect a promotion or a raise. We also strive to improve the accuracy of our predictions. We classify people into sub-classes because doing so allows us to set more accurate and precise expectations of them. For example, we categorize students on the basis of intelligence and motivation so that we can more accurately predict their chances of academic success. Classifying people according to age and general conditions of their health permits a closer estimation of their life expectancy. In fact, aptitude, readiness, guidance and personnel work are all predicated on the concept of prediction. Astrology, palmistry, and graphology are similar attempts at prediction.

Such expectations invariably are based on probability of occurrence for, as Fattu emphasizes, the only 100 percent accurate method of prediction is hindsight, an activity which is almost as popular as it is precise.[1]

[1] Nicholas A. Fattu, "Prediction: From Oracle to Automation," *Phi Delta Kappan,* 39 (June 1958) : 409–12.

BASES OF PREDICTION

The bases on which predictions are made range all the way from intuition and charlatanism to relatively precise empirically and theoretically derived relationships. Fortune-telling, for example, is simply guesswork made true by selective forgetting. Astronomy, on the other hand, is able to predict the position of a star at any given time in the future with relatively unlimited accuracy.

Prediction Based on Trends

Prediction is often based on trends. It appears safe, for example, to predict that the enrollment of our schools will increase in the next decade—it has been rising for some time now and probably will continue to do so for some time in the future. Trend prediction is a standard procedure in economics, where business conditions indexes, for instance, are based on the trend lines of relevant economic indicators. These are sometimes simply graphic extensions of the lines of growth of contributing factors. A more precise technique consists of deriving the mathematical equation of the line of growth and extrapolating it into the future. Some equations of this kind have relatively wide applicability: The Gompertz curve, for example, which displays an initial gradually increasing rate of acceleration, followed by a gradual tapering off to a limit, can be used to represent both population growth and the learning curve in the acquisition of a motor skill. The equation of the straight line is, of course, simpler. It is applicable to a number of situations, including perhaps changes in a phenomenon over a short period of time when growth can be assumed to be linear.

Prediction Based on Association

Prediction can also be based on the association between variables. For example, to the extent that IQ and academic achievement are positively correlated, it is possible to predict, with some degree of accuracy, that a person with a high IQ will probably do well academically. Such prediction is, of course, not free from exception, since the association between IQ and academic achievement is not a one-to-one relationship. Prediction on the basis of association can be extended to more sophisticated and precise predictions, involving formulas of various

degrees of complexity, the accuracy of which can be estimated through proper statistical techniques.

The most common basis for informal prediction used in education is performance on educational and psychological tests. By virtue of the correlation of the scores they yield with measures of certain other traits, all tests are in a sense prognostic. Intelligence tests, for example, are prognostic of academic and vocational success. Performance on an achievement test is predictive of later performance in the same area and, to a lesser degree, of performance in related areas. Every test allows for prediction and tests are rarely given for purposes other than prediction. This is implied in the concept of test validity.

At an elementary level, prediction can be based on a simple charting of two variables, such as IQ and grades, on a scattergram. A line of *best fit* can then be drawn through the data to display the general trend. This idea can be extended to provide an expectancy table or chart showing expected performance for different IQ levels. Table 13–1, for example, lists the

TABLE 13–1

Expected Achievement of First-Semester Freshmen[2]

OSU Scores	Probability of Earning a Point-Hour Ratio of at least				
	1.00	1.50	2.00	2.50	3.00
114–150	100	99	93	80	56
102–13	100	96	91	60	30
92–101	100	95	90	60	29
83–91	99	90	78	41	27
85–82	98	87	74	25	13
66–74	97	80	62	25	13
56–65	96	79	61	17	5
48–65	95	75	47	13	4
39–47	95	63	33	7	2
0–38	87	58	29	3	1

[2] Source: Bingham based on data from G. B. Paulsen. Walter V. Bingham, "Expectancies," *Educational and Psychological Measurement*, 13 (Spring 1953): 47–53.

probability of getting grade-point averages of 1.00, 1.50 and so on for freshmen of different levels of ability as measured by the Ohio State Psychological Examination.

Correlation

The concept of correlation is fundamental to prediction based on association among variables. This concept is adequately treated in introductory texts in statistics, and the discussion here will be restricted to a brief overview of its use in predictive studies. It must first be realized that correlation is not synonymous with causation; correlation simply implies concomitance. It may suggest causation, but the latter would have to be shown since, frequently, the correlation between factors is nothing more than the reflection of the operation of a third factor. For instance, there is a positive correlation between the number of churches and the number of traffic accidents in a given community, a relationship which can be explained readily on the basis of the growth in population to which both factors are related.

To be valuable in prediction, the degree of association between two variables must be relatively substantial, and, of course, the greater the association, the more accurate the prediction it permits. What this means in practice, however, is not clear, except perhaps that anything less than perfect correlation between two variables will permit errors in predicting one from a knowledge of the other. The correlation must, of course, represent a real relationship rather than simply the operation of chance.[3] Beyond this, however, what constitutes an adequate correlation between two variables can be appraised only on the basis of what can logically be expected, and, of course, what accuracy of prediction is required to serve the purpose of the study. A coefficient of correlation of 0.35 between motivation and grades, for example, is perhaps all that can be expected in view of the crudeness of our present measures of motivation and of grades.

Correlation is a group concept, a generalized measure which is useful primarily in predicting group performance. We can, for example, predict that gifted children as a group will succeed in school, but we cannot be sure that a particular

[3] A minimum value for a *coefficient of correlation* to be considered significant (as opposed to possibly being a chance relationship) can be estimated crudely at two to two-and-a-half times the reciprocal of the square root of the number of cases on which it was derived. The statistician will recognize these values as a rough estimate of the 5 percent and the 1 percent levels of significance of r.

gifted child will do well. Except where the correlation between two variables is ± 1.00, prediction always involves an element of risk, the magnitude of which can be appreciated from a consideration of the index of forecasting efficiency (e) which is defined as follows:

$$e = 1 - \sqrt{1 - r^2}$$

where r is the coefficient of correlation between the two variables. This measure represents the predictability of a coefficient of correlation over and above pure chance. Thus, using as an example, a coefficient of 0.60 between X and Y, we find

$$\begin{aligned} e &= 1 - \sqrt{1 - .3600} \\ &= 1 - \sqrt{.6400} \\ &= 1 - .80 \\ &= 0.20 \text{ or } 20\% \end{aligned}$$

On the basis of a correlation of 0.60 between X and Y, we can predict Y from a knowledge of X, 20 percent better than chance —that is, we can reduce by 20 percent the range of error that would be involved if we had based our prediction on pure guess.

Similar calculations of the predictive efficiency of other values of the coefficient of correlation would give the following:

r	e
± 1.00	100%
± 0.80	40%
± 0.60	20%
± 0.40	8%
± 0.20	2%
± 0.00	0%

Since most of the correlations among the variables of interest in education are of the order of 0.50, relatively little confidence can be placed in such predictions in the individual case. It is, therefore, necessary to attempt to raise the correlation, on the basis of which predictions are made, in order to increase their precision. This can be done by refining, either or both, the instruments used and the criterion being predicted,

and, as we shall see, by combining a number of variables into a composite predictor of the criterion.

STATISTICAL PREDICTION[4]

Simple Regression

The usual procedure for predicting one variable from knowledge of another consists of converting the correlation between the two variables into a predictive or *regression* equation, which expresses the relationship between them. Such an equation is simply the algebraic expression of the line of best fit through the data arranged in a scattergram on the X and Y co-ordinates. It gives the most probable value of the *dependent* variable (or *criterion*) for each of the values of the *independent* variable (or *predictor*).

The technique calls for the solution of the constants b and a in the equation of the straight line,

$$Y = bX + a$$

where the constants b and a simply indicate the slope and the "starting point" of the line of best fit.[5] This can be done directly either through the solution of the equations,

[4] This section will introduce the reader to simple mathematical reasoning. The concepts are not difficult; anyone who has had a course in elementary statistics should make it a point to follow the discussion. However, failure to understand the specific steps should not be a deterrent to grasping the general orientation of the presentation. Textbooks in statistics or in educational tests and measurements should be consulted for a more complete treatment.

[5] A common example of such a straight line is the formula, $F = 1.8\,C° + 32°$, from which Centigrade temperatures can be converted into Fahrenheit temperatures. The meaning of b and a are shown graphically in the chart in Figure 13-1.

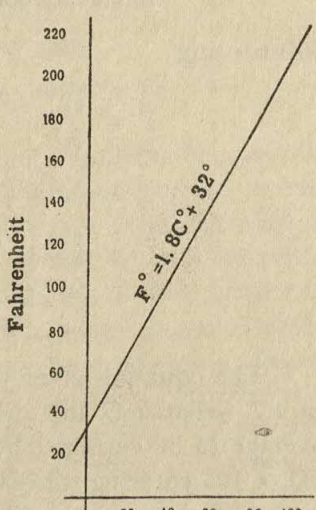


Figure 13-1

$$(1) \quad \Sigma Y = b\Sigma X + na$$
$$(2) \quad \Sigma XY = b\Sigma X^2 + n\Sigma X$$

or from the relationship,

$$\tilde{Y} = r\frac{\sigma_y}{\sigma_x}X + (Y - r\frac{\sigma_y}{\sigma_x}X)$$

where $\tilde{Y}$ is the estimate of the variable to be predicted; $\bar{X}$ and $\bar{Y}$ are the means, and σ_x and σ_y are the standard deviations of the variables X and Y. r is the coefficient of correlation between X and Y.

Thus, given:

$\tilde{Y}$ = John's estimated grade point average

X = John's IQ = 103

$\bar{Y}$ = School's average G.P.A. = 23

$\bar{X}$ = School's average IQ = 122

σ_y = Standard deviation of the school's distribution of G.P.A. = 0.7

σ_y = Standard deviation of the school's IQ = 11

r = the correlation between IQ and G.P.A. at this school = .50

Substituting

$$\tilde{Y} = r\frac{\sigma_y}{\sigma_x}X + (Y - r\frac{\sigma_y}{\sigma_x}X)$$
$$= .50\ (\frac{.7}{11})\ X + (2.3 - .50)\ (\frac{.7}{11})\ (122)$$
$$= .032X + (2.3 - 3.9)$$
$$= .032X - 1.6$$
$$= 3.3 - 1.6 = 1.7$$

The equation gives John a predicted grade-point average of 1.7. What does this mean? Actually, 1.7 is the grade-point average to be expected by the multitude of "Johns" with an IQ of 103 entering this particular college. It is simply an average to be expected from this hypothetical group—some will get more; some will get less. The next consideration is the variability to be expected in the grade-point average of this multi-

tude of students of IQ of 103. This introduces the concept of the *standard error of estimate,* or *prediction,* which can be calculated as follows:

$$\begin{aligned} S.E._{est.} &= \sigma_y\sqrt{1 - r^2} \\ &= 0.7\sqrt{1 - .2500} \\ &= 0.7\sqrt{.7500} \\ &= 0.7(.87) = 0.609 \text{ (or 0.6 approximately)} \end{aligned}$$

This can be interpreted on the basis of the normal probability distribution shown in Figure 13–2. Whereas the average

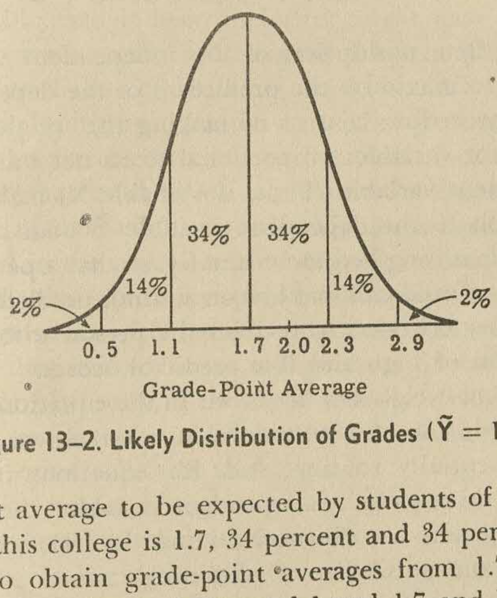


Figure 13–2. Likely Distribution of Grades ($\tilde{Y} = 1.7$)

grade-point average to be expected by students of this caliber attending this college is 1.7, 34 percent and 34 percent can be expected to obtain grade-point averages from 1.7 – 0.6 and from 1.7 + 0.6—that is, between 1.1 and 1.7 and between 1.7 and 2.3, respectively. Similarly, 14 percent of these students can be expected to obtain grade-point averages from 0.5 and 1.1 and between 2.3 and 2.9. And, some 2 percent can be expected to get a grade-point average of less than 0.5, and another 2 percent (approximately) a grade-point average in excess of 2.9.

One can go further and note that, since graduation generally requires a grade-point average of 2.0, only some 30 percent of these students might be expected to attain this level of scholarship. It must be remembered that the odds refer to the group and have only indirect meaning for any one student.

Multiple Regression

A more complex—and more realistic and useful—approach to the prediction of a given variable is *multiple regression*, in which the estimate of the criterion measure is based on the linear combination of several independent variables. For example, grades in college might be predicted from a linear combination of such variables as high-school rank (X_1); Scholastic aptitude score (X_2); and Cooperative English Test score (X_3) through an equation of the form,

$$\tilde{Y} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + a$$

where the Beta multipliers of the independent variables are computed to maximize the prediction of the dependent variable. The procedure consists of making the weight of each of the predictor variables proportional to its net contribution to the dependent variable. Thus, if Variable X_1 makes a greater contribution to the dependent variable $\tilde{Y}$ than does a lesser predictor, it is weighted more heavily, so that a person high on an important predictor and low on a minor predictor would obtain a higher criterion score than the person with the reverse combination of high and low predictor scores.

The Beta-weights to be placed in the equation are derived from the solution of *n* equations in *n* unknowns. The procedure is essentially routine, and, for equations involving no more than three or four independent variables, the work is not particularly time- or effort-consuming. In fact, with the computer and the possibility of obtaining canned programs, the work is now essentially clerical. The computation is beyond the scope of this text, but it is not beyond the capabilities of any teacher willing to solve simultaneous equations.

Once the weights have been determined, an estimate of an applicant's likely success can be obtained by the simple process of substituting his scores on each of the independent variables in the equation. Again, this does not represent exactly what the student will get, but an average predicted by the equation for the multitude of applicants whose weighted summation of X_1, X_2, and X_3 gives such a predicted score. The procedure also permits the computation of the probability of a given student

achieving any given grade-point average above or below that predicted.

1. *Choice of predictors.* A regression equation is only as good as the variables on which it is based. The specific combination of variables to be included in the equation, and the relative weights they are to be given, depends first of all on the nature of the problem; different variables would be needed, for instance, for predicting success in engineering than in education. They would also vary from subject to subject, from school to school (even in the same subject), and from year to year. Furthermore, in the present stage of development of the social sciences, one can never be sure of what to include, for some excluded variable, had it been included, might have tapped an important aspect of performance ignored by the other components of the predictive equation. It would be desirable, for instance, in developing an equation to predict performance in a relatively unknown area to err on the side of including too many variables rather than too few, especially since those variables that do not contribute significantly to the prediction will be eliminated in the process of deriving the equation.

The variables selected as predictors should be correlated as highly as possible with the dependent variable, but, on the other hand, they should be as independent of each other as possible, since, obviously, if two variables are duplicates of one another, one or the other will carry the prediction all by itself, and the other will add nothing.

A common misuse of regression equations is to include an unnecessarily large number of variables on a trial-and-error basis, thus increasing their complexity and unwieldiness. Although this may be permissible, and even advisable, in exploratory studies, the investigator should generally attempt to select predictors that fit into the theoretical structure of the variable to be predicted. Maximum effectiveness in prediction would tend to be obtained when factorially pure predictors are selected to cover each of the components of the criterion, with a maximum of validity and a minimum of overlapping.

The variables included also have to be capable of reliable measurement and be available before the prediction is needed. The variables used in predicting college success, for example, have to be available at the time the student applies for admis-

sion. Yet, one must be careful not to select variables simply because they are available or easy to measure. Most predictive equations, for example, do not place sufficient emphasis on such variables as motivation and personality characteristics, principally because they are difficult to measure with sufficient precision for them to add significantly to the equation.

An interesting question in the use of multiple regression equations in personnel selection concerns the role that the judgment of the personnel interviewer is to play in the selection. Although sizable individual differences are bound to exist in the ability of interviewers to distinguish between potentially satisfactory and potentially unsatisfactory workers, there is sufficient evidence to suggest the relative inadequacy of personal judgment to warrant caution in its use. The problem can perhaps be resolved by having the interviewer's rating of the applicant included as one of the predictors in the equation. If the rating has validity with respect to the criterion scores over and above what is already contributed by the other variables in the equation, its Beta-weight will testify to its usefulness, and the rating should be kept either as part of the equation or as a separate step in the hiring process. If, on the other hand, nothing is added by the inclusion of the interview rating, employment procedures could be streamlined by the omission of the interview without loss in predictive efficiency.

2. *Choice of criterion.* The choice of the criterion of a regression equation is a vital factor in its effectiveness. It is an even more crucial factor in its validity for, obviously, what constitutes an adequate criterion depends on the purpose of the study. More specifically, the criterion should reflect the objectives of what is to be promoted. In practice the dependent variable selected is too frequently a relatively inadequate criterion of the true goals of the activity in question. Generally, the criterion being predicted in schools of education, for example, is the grade-point average which, besides incorporating a considerable element of unreliability and invalidity as a measure of learning, is in itself only vaguely related to the crucial question "Will the student make a good teacher?" Similarly, it is conceivable that a predictive equation used in medical-school admissions will select scholars, and not necessarily promising physicians.

Even when grades are a legitimate criterion of college success, a number of questions still need to be answered: "What grades are to be counted: grades for the first semester, the first year, or the four years? Should we include all subjects, or should we include only grades in courses in the immediate area of the student's academic goal? and so on. To the extent that a criterion is unclear, it is relatively difficult to devise an equation for its adequate prediction. Before grades can be used as a criterion, for example, it is necessary for the teachers to synchronize their grading procedures from the standpoint of validity and reliability and—when a number of teachers are involved—from the standpoint of calibration to a common point of reference. Predicting freshman grades in college, for example, is complicated by the fact that the different schools within a university enroll students in different courses, each with its own emphasis and its own grading standards. For this reason, it is generally necessary to prepare separate equations for the different schools, except where a general college has charge of the total lower division of the university program.

A special problem with respect to the use of grades as a criterion is generally encountered in the graduate school, where "everybody gets a B." Because of the restriction in the range over which graduate grades are assigned, the correlation of the various predictors with grades as criterion is automatically negligible. In addition, the predictors—for example, undergraduate record, aptitude, Graduate Record Examination scores, and so on—are generally highly inter-correlated. Before adequate prediction can be obtained, therefore, it is necessary to increase the range over which graduate success is measured and to seek predictors of greater mutual independence.

3. *Shrinkage.* A phenomenon peculiar to regression equations derived for the purpose of predicting a given criterion is that, when the formula is applied to any group other than the one on which its Beta-weights were obtained, there is a shrinkage in its degree of accuracy. This might be anticipated from the premises on which the equation is derived. Since the Beta-weights are obtained to maximize the prediction of the dependent variables in the particular group under study, even to the point of capitalizing on chance factors, it follows that for any other group the predictors will not fit as well, and that the

overall prediction will shrink somewhat from this ideal level.

It has been suggested that, in predicting a given criterion from certain independent variables, two equations be derived from different samples, and that the weights obtained in these samples be averaged as the best estimate of the prediction in the overall population, freed from the idiosyncrasies of either of the particular samples. This is probably sound, since such an averaging would stabilize the Beta-weights and would provide for a more adequate prediction in the general case. Two implications are involved here: 1. a slight change in the Beta-weights of the predictors does not affect the prediction appreciably and is not particularly objectionable from a theoretical point of view, and 2. any regression equation should always be cross-validated on a second sample. There is also the need for periodic revisions of the equation.

Other Forms of Prediction

1. *Non-linear regression.* Most predictions in education are based on the assumption of a linear relationship among variables, an assumption which tends to be fulfilled in most of the variables of interest to educators. There are, however, instances in which the variables are not linearly correlated, and it may be necessary either to transform them to a new scale or, perhaps, to consider a non-linear equation of a form such as:

$$\tilde{Y} = \beta_1 X_1 + \beta_2 X^2_2 + \beta_3 X^3_3 + a$$

2. *Pattern analysis.* An even more advanced prediction technique is pattern analysis (or *profile* or *configurational* analysis), which is generally considered more accurate than the traditional multiple regression technique when dealing with certain variables. This technique is, of course, beyond the scope of the present text.

3. *Discriminant functions.* It frequently occurs that the phenomenon to be predicted is qualitative rather than quantitative. It may be desirable, for example, to allocate freshmen to advanced, average, and introductory classes on the basis of such independent variables as IQ, high-school rank, and previous background in the subject. The problem is to weight the predictor variables so that the distinction between the categories into which the subjects are to be assigned is maximized.

This can be done through discriminant analysis, a technique, devised by Fisher,[6] which attempts to set up the linear combination of weighted measurements that will maximize the discrimination between groups with a minimum of overlapping and false classification. More specifically, the technique attempts to derive a weighted linear function of a set of variables that will maximize both the in-group homogeneity and the between-group distinction.

Evaluation of Multiple Regression

Multiple regression equations are obviously practical. Despite the shrinkage that accompanies their application to a different group, the increase in correlation between predictors and criterion, which multiple regression provides, frequently results in a considerable increase in accuracy of prediction.

On the other hand, predictive studies are essentially, if not entirely, empirical, and their contribution to the development of education as a science is relatively limited. Correlations, no matter how weighted, add little to the development of science, and, though hypotheses and theory may be involved in the selection of the variables to be included in the equation, thus far the procedure has made relatively little attempt at the discovery of the fundamental relationships among phenomena. Thus the review of over one thousand studies attempting to relate one or more tests to the prediction of some aspect of academic achievement led Travers to conclude that the actual contribution to knowledge made by these studies is relatively small.[7]

Frequently any number of variables are thrown together in the equation, and reliance is placed on the statistical procedure to eliminate variables which make only a minor contribution to the prediction of the criterion. With the increasing availability of the computer, this trial-and-error process may become more prevalent. It may add to the accuracy of prediction and, especially, to the inclusion of variables that would otherwise be overlooked, but it will provide relatively little insight into the basic reasons underlying their contribution. Including a whole

[6] Ronald A. Fisher, *The Design of Experiments* (6th ed.; New York: Hafner, 1951).

[7] Robert M. W. Travers, "The Prediction of Achievement," *School and Society*, 70 (November 1949): 293–4.

slough of variables that might conceivably be useful increases the work of deriving the equation and, of course, of using it after it has been derived. In the meantime, it obviates the need for theory to guide the investigation and makes the process clerical rather than scholarly.

An interesting phenomenon of prediction is the tendency for the prediction to be either self-fulfilling or self-destroying in that it stimulates a reaction which interferes with its fulfillment. For example, if an applicant for college admission is told that he is a borderline case, he is likely to exert himself a little more than he would normally and thus invalidate the prediction. Or, if, because his chances for success are slim, he is advised to carry a light load, his grades may be considerably above the level predicted.

SPECIAL CASES

Individual Prediction

Probability as it underlies prediction, is a group concept whose applicability to the individual needs to be considered. Prediction in the individual case can be based either directly on the individual's past performance or on the assumption that the probability which pertains to the group of which he is a member somehow has a bearing on him. If, for example, we can determine by extended observation that the individual plays golf a couple times a month, we might establish the probability of his playing in any one week to be 0.50. Such a prediction would, of course, be subject to the usual risk of error.

Individual prediction based on group probability is somewhat more theoretically complex. To say that an entering freshman has one chance in four of passing a certain test is a group concept which technically does not apply to the individual. What is said is that some 25 percent of the group will succeed. As it applies to the individual, however, the probability statement is in the form of a dichotomy; he will not be 25 percent successful, but will pass or fail. Both the gifted and the dull student will either succeed or fail on a given test, each with the identical probability level of one or nothing; it still seems relevant, however, to note that the underlying group probabilities of success may be 0.98 and 0.03 respectively.

Clinical versus Actuarial Prediction

The discussion so far has emphasized predictions based on the formal weighting of the variables in some form of a predictive equation. Predictions can also be based on a *judgmental* or *clinical* approach.

A fundamental aspect of clinical prediction involves the categorizing of phenomena into homogeneous sub-classifications, each with a definite probability expectancy with respect to certain outcomes. Clinical predictions of longevity, for instance, might involve categorizing people into finer and finer sub-classifications—not human beings but Americans; not Americans, but American men. Further sub-classifications can be made by age, marital status, occupational status, and so on. To the extent that the individual is pigeon-holed correctly into a highly homogeneous sub-classification for which group probabilities relative to outcome have been determined, a relatively accurate prediction can be made.

In practice, it is impossible to devise such classifications on the basis of all factors related to a given phenomenon. Furthermore, even if it were possible to devise the categories and to determine the probability to be attached to each, it would be impossible to assign the individual to the particular sub-category representing each of his particular traits with any degree of certainty. The categorizing of the individual must be made, therefore, on the major bases of classification only, and, for that reason, such predictions—like all predictions—are only probable.

The relative effectiveness of actuarial (statistical) and clinical predictions has received considerable attention since the publication by Meehl[8] of a booklet on the subject. The evidence reviewed by Meehl and others tends to favor actuarial prediction. On the other hand, care must be taken not to overgeneralize; a more tenable position might be that each approach is probably the more accurate for certain types of problems, and that a clinical psychologist, for example, will always have to base certain prognoses on a clinical rather than on a

[8] Paul E. Meehl, *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* (Minneapolis: University of Minnesota Press, 1954).

statistical foundation. It might be said that statistical prediction is perhaps generally more accurate in cases involving statistical relationships, and that clinical predictions are more effective in cases involving clinical data of a qualitative and judgmental nature. There is, of course, much good sense in Stouffer's suggestion that the statistician and the clinical worker would both gain if they would stop quarreling with each other and begin borrowing what each has to contribute.[9]

Prediction of rare events. A unique aspect of prediction concerns the case in which the phenomenon to be predicted has a very low probability of occurrence—for example, death as a result of an abcessed tooth. Meehl and Rosen suggest that in such cases a greater overall accuracy of prediction would be obtained by predicting for each person that he will not meet death in this way than by attempting to predict which person will and which person will not die in this way.[10]

Prediction and Determinism

Another interesting point concerning prediction is discussed by Feigl and Brodbeck,[11] who raise the question of whether our ability to predict delinquency would absolve the individual of any blame when he becomes a delinquent according to prediction; or is he robbed of any credit when he refrains from becoming delinquent when the equation predicts such an outcome? This point borders upon fatalism and determinism and is an interesting point in the philosophy of science. The reader is referred to the original source for further discussion.

SUMMARY

1. Man is perpetually predicting the likely outcomes of his efforts. Basic to all such predictions is the element of risk at various levels of probability.

2. Prediction can be based on trends derived from past performance. As long as the same forces continue to act on a given

[9] Samuel A. Stouffer, "Notes on the Case-Study and the Unique Case," *Sociometry,* 4 (November 1941) : 349–57.

[10] Paul E. Meehl and Albert Rosen, "Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores," *Psychological Bulletin,* 52 (May 1955) : 194–216.

[11] Herbert Feigl and May Brodbeck (eds.), *Readings in the Philosophy of Science* (New York: Appleton-Century-Crofts, 1953).

object or factor, it seems logical to expect that it will continue in the same general direction at the same rate of speed or acceleration.

3. Prediction can also be based on association between variables as represented by the concept of correlation. At its most elementary level, prediction can be a matter of general expectancy of superior, average, or inferior status on one factor, by virtue of its status on a related variable. A more precise prediction can be obtained by converting the measure of association into a regression equation involving one or more independent variables. Such an approach also provides an estimate of the margin of error in the prediction.

4. The choice of the variables (both predictor and criterion) is of primary importance in the adequacy of the equation derived. In a multiple regression equation, for example, the predictor variables should be as highly correlated with the criterion and as independent of one another as possible. Generally, the choice of predictors should make logical and theoretical—rather than simply empirical—sense. It is sometimes possible to cluster variables in order to prevent undue unwieldiness in the equation.

5. While statistical prediction generally consists of the prediction of a criterion score through the weighted linear combination of relevant predictor variables, prediction can also be based on a non-linear combination of variables. Prediction in discriminant analysis is oriented toward the optimal allocation of individuals to different groups rather than the derivation of a predicted criterion score.

6. Statistical prediction, thus far, has been empirically—if not, clerically—oriented and has contributed little to the advancement of education as a science. Greater emphasis needs to be placed on the theoretical considerations underlying the relationships on which the prediction is based.

7. Prediction is a group concept that has only indirect meaning and application to the individual.

8. Of considerable current interest is the relative effectiveness of clinical and actuarial prediction. Clinical prediction is based on the concept of the classification of the individual into relatively homogeneous sub-classes for which a relatively high level of group probability has already been established. While it is difficult to generalize, it is likely that each approach is the more valid in situations for which it seems most appropriate. However, both approaches are complementary rather than antagonistic.

PROJECTS and QUESTIONS

1. What procedures are currently used in your college for the admission of students? How satisfactory have present screening procedures been? With the use of the computer (if available) devise a predictive equation to predict the relative achievement of

applicants. Choose the variables to be incorporated into the equation carefully on the basis of both practical and theoretical considerations. Check on the validity of your equation by relating predicted scores to actual performance at the end of the semester or year.

2. Make a brief survey of employee-selection procedures of local industrial firms. What evidence do they collect systematically of the effectiveness of their procedures?

SELECTED REFERENCES

ADRAGNA, C. MICHAEL. "Prediction of Achievement in Junior High School General Science," *Dissertation Abstracts,* 14 (December 1954): 2289.

BERDIE, RALPH F. and SUTTER, NANCY A. "Predicting Success of Engineering Students," *Journal of Educational Psychology,* 41 (March 1950): 184–90.

BROWN, GEORGE W. "Discriminant Function," *Annals of Mathematical Statistics,* 18 (December 1947): 514–28.

BRYAN, JOSEPH G. "The Generalized Discriminant Function: Mathematical Foundation and Computational Routine," *Harvard Educational Review,* 21 (No. 2, 1951): 90–5.

CAPPS, MARIAN P. and DECOSTA, FRANK A. "Contributions of the Graduate Record Examinations and the National Teachers Examinations to the Prediction of Graduate School Success," *Journal of Educational Research,* 50 (January 1957): 383–9.

CARLILE, AMOS B. "Predicting Performance in the Teaching Profession," *Journal of Educational Research,* 47 (May 1954): 641–68.

COTTRELL, LEONARD S. "The Case-Study Method in Prediction," *Sociometry,* 4 (November 1941): 358–70.

EDUCATIONAL TESTING SERVICE. "Clinical and Actuarial Prediction in a Setting of Action Research," in *1955 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1955.

EZEKIEL, MORDECAI and FOX, KARL A. *Methods of Correlation and Regression Analysis.* New York: Wiley, 1959.

FATTU, NICHOLAS A. "Prediction: From Oracle to Automation," *Phi Delta Kappan,* 39 (June 1958): 409–12.

FEIGL, HERBERT and BRODBECK, MAY (eds.). *Readings in the Philosophy of Science.* New York: Appleton-Century-Crofts, 1953.

FREDERIKSEN, NORMAN O. "Predicting Mathematics Grades of Veteran and Non-Veteran Students," *Educational and Psychological Measurement,* 9 (Spring 1949): 73–88.

FRICK, JAMES E. and KEENER, HELEN E. "A Validation Study of the Prediction of College Achievement," *Journal of Applied Psychology,* 40 (August 1956): 251–2.

GARRETT, HARLEY F. "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts

and Science and Teachers Colleges," *Journal of Experimental Education,* 18 (December 1949): 91–138.

HOLZBERG, JULES D. "The Clinical and Scientific Methods: Synthesis or Antithesis," *Journal of Projective Techniques,* 21 (September 1957): 227–42.

HORST, PAUL (ed.). "The Prediction of Personal Adjustment," *Social Science Research Council Bulletin,* 48 (1941): 1–156.

———. "A Technique for the Development of a Differential Prediction Battery," *Psychological Monograph,* 68, No. 9, 1954.

HORST, PAUL and MACEWAN, CHARLOTTE. "Optimal Test Length for Multiple Prediction: The General Case," *Psychometrika,* 22 (December 1957): 311–24.

HOTELLING, HAROLD. "Problems in Prediction," *American Journal of Sociology,* 48 (July 1942): 61–76.

JACKSON, ROBERT A. "Prediction of the Academic Success of College Freshmen," *Journal of Educational Psychology,* 46 (May 1955): 296–301.

JENSEN, RALPH E. "Predicting Scholastic Achievement of First-Year Graduate Students," *Educational and Psychological Measurement,* 13 (Summer 1953): 322–9.

JOHNSON, PALMER O. "The Quantification of Qualitative Data in Discriminant Analysis," *Journal of the American Statistical Association,* 45 (March 1950): 65–76.

———. "The Best Linear Estimate of the Predicted Value and the Standard Error of the Estimate," *Journal of Experimental Education,* 25 (March 1957): 233–9.

KACZKOWSKI, HENRY R. and ROTHNEY, JOHN W. M. "Discriminant Analysis in Evaluation of Counseling," *Personnel and Guidance Journal,* 35 (December 1956): 321–5.

KING, RICHARD G. *The Prediction of Choice of Undergraduate Field of Concentration in Harvard College through Multiple Discriminant Analysis.* Cambridge: Harvard University Press, 1958.

KLEIN, GEORGE S. "An Application of the Multiple Regression Principle to Clinical Prediction," *Journal of General Psychology,* 38 (April 1948): 159–79.

KVARACEUS, WILLIAM C. "Prediction of Maladjustive Behavior," in *1958 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1958.

LANNHOLM, GERALD V. and SCHRADER, WILLIAM B. *Predicting Graduate School Success: An Evaluation of the Effectiveness of the Graduate Record Examinations.* Princeton: Educational Testing Service, 1951.

LARSON, SEDMER C. "The Shrinkage of the Coefficient of Multiple Correlation," *Journal of Educational Psychology,* 22 (January 1931): 45–55.

LUBIN, ARDIE. "Linear and Non-Linear Discriminant Functions,"

British Journal of Psychology; Statistical Section, 3 (June 1950): 9–104.

LUNDBERG, GEORGE A. "Case Studies vs. Statistical Methods: An Issue Based on Misunderstanding," *Sociometry,* 4 (November 1941): 379–83.

MCHUGH, RICHARD B. and APOSTOLAKOS, PETER C. "Methodology for the Comparison of Clinical with Actuarial Predictions," *Psychological Bulletin,* 56 (July 1959): 301–8.

MEEHL, PAUL E. *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis: University of Minnesota Press, 1954.

———. "When Shall We Use our Heads Instead of the Formula?" *Journal of Counseling Psychology,* 4 (Winter 1957): 268–73.

MEEHL, PAUL E., *et al.* "Symposium on Clinical and Statistical Prediction," *Journal of Counseling Psychology,* 3 (Fall 1956): 163–73.

UNIVERSITY OF MINNESOTA. *University of Minnesota Studies in Predicting Scholastic Achievement.* Minneapolis: University of Minnesota Press, 1942.

PICKREL, EVAN W. "Classification Theory and Techniques," *Educational and Psychological Measurement,* 18 (Spring 1958): 37–46.

ROGERS, CARL R. "Implications of Recent Advances in Prediction and Control of Behavior," *Teachers College Record,* 57 (February 1956): 316–22.

RULON, PHILLIP J. "Distinctions between Discriminant and Regression Analysis and a Geometric Interpretation of the Discriminant Function," *Harvard Educational Review,* 21 (Spring 1951): 80–90.

SARBIN, THEODORE R. "The Logic of Prediction in Psychology," *Psychological Review,* 51 (July 1944): 210–29.

SEVIER, FRANCIS A. C. "Testing the Assumptions Underlying Multiple Regression," *Journal of Experimental Education,* 25 (June 1957): 323–30.

SHEA, JOSEPH A. *The Predictive Value of Various Combinations of Standardized Tests and Subtests for Prognosis of Teaching Efficiency.* Educational Research Monograph, 19, No. 5. Washington, D.C.: Catholic University of America Press, 1955.

STOUFFER, SAMUEL, *et al. Measurement and Prediction: Studies in Social Psychology in World War II.* Princeton: Princeton University Press, 1950.

STUITT, DEWEY B., *et al. Predicting Success in Professional School.* Washington, D.C.: American Council on Education, 1949.

TIEDEMAN, DAVID V. and BRYAN, JOSEPH. "Prediction of College Field of Concentration," *Harvard Educational Review,* 24 (Spring 1954): 122–39.

TIEDEMAN, DAVID V., *et al.* "The Multiple Discriminant Function—

A Symposium," *Harvard Educational Review,* 21 (Spring 1951): 71–95.

TRAVERS, ROBERT M. W. "The Prediction of Achievement," *School and Society,* 70 (November 5, 1949): 293–4.

———. "Significant Research on the Prediction of Academic Success," in Donahue, Wilma R. (ed.). *The Measurement of Student Adjustment and Achievement.* Ann Arbor: University of Michigan, 1949.

WALLACE, WIMBURN L. "The Prediction of Grades in Specific College Courses," *Journal of Educational Research,* 44 (April 1951): 587–97.

WALLIN, PAUL. "The Prediction of Individual Behavior from Case Studies," in Horst, Paul. *The Prediction of Personal Adjustment.* New York: Social Science Research Council Bulletin, 48, 1941.

Just as some men prefer to grasp a crude hoe, bend the back and chop away, some people desire to live academically simple lives. Too many teachers attempt to pedal around the world of knowledge in an intellectual unicycle.

JOHN B. BARNES

14 Educational Research: A Review and Evaluation

PRESENT STATUS OF EDUCATION

Keynote on Progress

The advances we are witnessing in the various fields of scientific endeavor are truly spectacular. Not only have we been able to accomplish feats that only a few years ago belonged in the realm of science fiction, but our achievements will undoubtedly become progressively more impressive. Of particular significance in this progress is the almost complete orientation of modern industry toward research as the foundation on which companies—and even nations—must depend for progress and, indeed, for survival. Those who did not experiment were left manufacturing buggy-whips and are no longer, while companies with a progressive and imaginative management have blossomed from backyard garages into national and international prominence. This is the story of America on the industrial and research front. Where new products can do jobs better, where sleeker cars can be sold, where wars need to be won, industrial reliance on research can be depended on.

Progress in Education

Progress in education has been far less impressive, and the question is whether the students we are now teaching can be developed into people big enough to live in this rapidly moving world. We have made some gains—in fact, we have made very definite gains under relatively adverse circumstances. Over the last few decades, we have attained a much better understanding of the child as a developing organism and as a learner, of his learning process, and of the role of education in promoting his maximum growth. We have made great strides in viewing the curriculum from the standpoint of the pupil, and in adapting teaching methods and classroom organization to the task of promoting the maximum self-realization of the pupil. We are much more conscious of the effects of experience on the growth and welfare of the child and of the methods by which they can be promoted most effectively. We have made gains in the area of motivation, and have made substantial progress in modifying our curriculum and our teaching methods to capitalize on the pupil's interests, needs, goals, and purposes.

We have made notable advances in research methods. We have certainly come a long way from the crude work of Binet in testing[1] and Rice in experimentation.[2] We have made special gains in certain areas—for example, in reading, where yearly reviews[3] present an impressive list of accomplishments. We are no longer in a sea of ignorance; we can now locate areas of knowledge, and we have a fair idea of those that still need exploration. We have even found that certain problems cannot be solved. Perhaps most important, we are orienting ourselves more toward research, and we have gained some understanding of the type of research on which educational gains must be based.

Every year, hundreds of studies are added to the large number already reported in professional literature. And, of

[1] Alfred Binet, see Chapter 15.
[2] Joseph M. Rice, see Chapter 15.
[3] William S. Gray, "Summary of Reading Investigations," *Journal of Educational Research*, February issue, yearly, 1924–1960. Continued, Helen M. Robinson, (1961), and Theodore L. Harris, (1962-date).

course, the *Encyclopedia of Educational Research* stands as an impressive monument to the research activities of the profession. Not all of these studies have been earth-shaking; on the contrary, many of them have been very elementary. In many areas, the evidence is still meager and fragmentary, and many problems are still relatively unexplored. But it must be remembered that education as a scientific undertaking dates back only to the early 1900's. We must remember that it has taken centuries to turn quackery into medicine, astrology into astronomy, and alchemy into chemistry.

If we trace the course of development of education as a science since 1900, we find that the growth has been largely in recent years. The period from 1900 to 1915 was a period of expansion during which the groundwork was established in a number of directions, particularly in statistics and testing. From 1915 through 1940, these procedures and techniques became more widespread, and a large number of investigations were conducted. Since the war, we have been more concerned with a critical analysis of our methodology and are just now setting the foundation for accelerated growth.

Early progress was particularly delayed by the relative lack of statistical techniques and of instruments of testing. These problems have been relatively alleviated in recent years, and our progress should be correspondingly easier and faster. Still very much with us as a retarding influence in the progress of education as a science, however, is the complexity of human nature, the nebulous and intangible nature of educational phenomena, and the time it takes for changes in these phenomena to take place. Educational progress has also been hampered by the highly decentralized nature of our educational system, which has minimized the interstimulation of the members of the profession and, of course, has led to considerable duplication in research.

Not only have we had to start from scratch in a field that does not lend itself easily to certain types of research, but we have had to overcome conditions of overcrowded classrooms, overworked and undertrained teachers, public indifference toward research, and traditionalism on the part of the teaching profession itself. Yet, despite these limitations, we have forged

ahead; Burton,[4] for example, provides a long list of educational gains which he feels can legitimately be attributed to research. The twenty-fifth anniversary edition of the *Review of Educational Research* (June 1956) also paints a rather impressive picture of growth in educational research.

Lag in Education as a Science

While many educators feel we owe no apologies for professional unproductivity, education is probably characterized more by gaps than by solutions, and, despite the many gains we have made, we cannot avoid agreeing with Kandel who, on looking at the *Encyclopedia of Educational Research,* wondered to what extent the mountain of material reviewed there would lead to improvement in educational practice.[5] An even more pessimistic view is that of Lamke who, in 1955, expressed the opinion that if the research in the previous three years in medicine, agriculture, physics, and chemistry were to be wiped out, our life would be changed materially, but if research in the area of teacher personnel in the same three years were to vanish, educators and education would continue much as usual.[6] Fehr points out that much of what is called educational research is not research at all when gauged by scientific standards.[7] Similarly, Tate feels that, outside of the very narrow field of psychometrics, the contribution of educational research has been wholly disappointing, that little of any scientific value can be derived from the "tons" of research that have been conducted, and that the majority of the studies are unreliable, trivial, and unworthy of serious consideration, much less application.[8] A similar view is presented by Eurich who suggests that unfortunately much educational research can be characterized as "the accumulation of irrelevant statistics in order to pro-

[4] William H. Burton, "Basic Principles in a Good Teaching-Learning Situation," *Phi Delta Kappan,* 39 (March 1958) : 242–8.

[5] Isaac L. Kandel, "Educational Research," *School and Society,* 72 (October 7, 1950) : 232.

[6] Tom A. Lamke (Chairman, Committee on Teacher Personnel), "Introduction," *Review of Educational Research,* 25 (June 1955) : 192.

[7] Howard F. Fehr, "Present Research in the Teaching of Arithmetic," *Teachers College Record,* 52 (October 1950) : 11–23.

[8] Merle W. Tate, "Operationalism, Research, and a Science of Education," *Harvard Educational Review,* 20 (January 1950) : 11–27.

ceed from an unwarranted hypothesis to a foregone conclusion."[9] Kerlinger points out that educational research is frequently characterized by triviality, superficiality, and scientific naïveté.[10]

We continue to make the same mistakes, to follow the same useless leads, and to clutter journals with studies that add little, if anything, to the development of education as a science or to the improvement of educational practice. For example, according to Charters, over and over again in the last thirty years, research studies have attempted to describe the social composition of school-board members. Every study "discovers" the same set of facts and, obviously, adds nothing to our understanding of educational problems. He points out that the frontiers of science have moved beyond the student who persists in surveys of school-board member characteristics, and that the educational researcher must seek new ways to answer new questions.[11] Although it is true that education as a science is relatively young, and that educational research frequently must be conducted under conditions seriously short of ideal, it must be recognized nonetheless that educational progress is being impeded by a number of weaknesses that need to be fully recognized and evaluated from the standpoint of possible improvement, if education is to forge ahead.

Lack of Orientation toward Research

Our failure to keep pace with the progress in other fields of scientific endeavor can be traced directly to our insufficient appreciation of the importance of research as the vehicle on which scientific progress must depend. While research is an integral part of business and industry, this relationship is still far from clear in education where the emphasis is on teaching rather than on a balance between teaching and research into what and how one should teach. As Stanley suggests, "no modern competitive business could survive long if it put as little

[9] Alvin C. Eurich, "New Dimensions in Educational Research," *A.E.R.A. Newsletter*, 12 (April 1962) : 4–10.

[10] Fred N. Kerlinger, "Mythology of Educational Research: The Methods Approach," *School and Society*, 88 (March 1960) : 149–51, 363–4.

[11] W. W. Charters, "Beyond the Survey in School Board Research," *Educational Administration and Supervision*, 41 (December 1955) : 449–52.

money, time, and effort into careful research and development as our public schools."[12]

The current lack of orientation of education toward science and research is illustrated by Remmers' interesting comparison of our outlook toward the material components of our culture with that toward the social aspects.

Our thinking is

Forward-looking in areas of material culture	Backward-looking in areas of social culture
Experimental Attitude: We view proposed changes without prejudice and with open minds, subordinating emotional considerations and demanding factual evidence upon which to base tentative working conclusions.	*Stand pat attitude:* We view proposed change with biased outlooks and strongly emotional convictions, subordinating rational considerations to cherished traditions, beliefs, and loyalties.
Old ideas held invalid: We are sure that in ten years or less many of our present theories and practices will be out of date, and will need to be discarded as obsolete.	*New ideas held unsound:* We are convinced that theories and practices of a century or more ago are patently infallible and should remain essentially unchanged forever.
The past viewed with amusement: We laugh at the scientific notions of the last generation.	*The future viewed with alarm:* We dread the social convictions of the next generation.
Change welcomed as progress: Having identified technological change with cultural progress, we acclaim each advanced material invention as a new Promise of American Life and Progress.	*Change opposed as regression:* Having identified social change with cultural decay, we denounce each advanced social invention as another Portent of American Decay and Death.[13]

[12] Julian C. Stanley, "Studying Status vs. Manipulating Variables," in Raymond O. Collier and Stanley M. Elam (eds.), *Research Design and Analysis* (Bloomington: Phi Delta Kappa, 1961), pp. 173–208.

[13] Hermann H. Remmers, "The Expanding Role of Research," *North Central Association Quarterly*, 23 (April 1949): 369–76.

Educators have made research an incidental and haphazard activity capable of providing only partial, if not erroneous, answers. As Brownell points out, our "shoestring" approach has exposed us to the ridicule of our more scientifically sophisticated colleagues.[14] In contrast to the physical sciences, where professional research workers devote their whole lives to the discovery of the principles of a small phase of a given field, education has allowed the amateur and the hobbyist to carry the ball. A study by Brownell of articles in professional journals revealed that 44 percent of the research studies reported in the field of arithmetic were done by individuals who apparently terminated their research efforts with one study. The corresponding figure for reading is 54 percent, for spelling 70 percent, and for English 63 percent. Or, stated differently, 79 percent of the arithmetic authors reported only one study; the corresponding figure for reading is 66 percent, for spelling 82 percent, for English 82 percent. Of the 778 authors of arithmetic articles, only 10.5 percent had published three or more articles. The corresponding figure for reading was 11.2 percent, for spelling 7.8 percent, for English 8 percent. In other words, three published reports qualifies research workers in education for membership in the highest 10 percent from the standpoint of productivity.[15]

Administrators are frequently adverse to the conduct of research; not only will they not conduct research themselves —especially pure research—but they also tend to refuse to give others permission to conduct research within their systems. To some degree, this reluctance to carrying out research is understandable since research frequently disturbs classroom routine and sometimes stirs parental objection. Administrators point out that their primary responsibility is to the children in the classroom; that the school exists for the purpose of teaching children and not for using them as guinea pigs, and that it is not legitimate to use taxpayers' money for research when present techniques are "satisfactory." This argument is

[14] William A. Brownell, "The Case for Educational Research," *Phi Delta Kappan*, 37 (February 1956) : 203–6.

[15] William A. Brownell, "A Critique of Research on Learning and on Instruction in the School," *Graduate Study in Education* (50th yearbook, National Society for the Study of Education, Pt. I, Chicago: University of Chicago Press, 1951), pp. 62–66.

sound as far as it goes but—even though the parallel is not complete—it might be pointed out that, if industry conceived its responsibility to its stockholders as simply selling its products and not experimenting, it would soon be out of business and have on hand a large supply of "buggy whips."

If one of the major functions of the school is to provide social leadership, it needs to accelerate its efforts toward solving the problems in its own house. Most of the major disciplines, especially the physical sciences, have placed strong emphasis on research—and with excellent results. It is high time for educators to realize the need for keeping pace with the world their efforts have helped produce. It is not expected that education have all its problems solved, or even in the process of solution, or that it rely exclusively on scientific procedures, for it is fully realized that there are significant aspects of education that do not lend themselves to scientific determination. There is, however, a need for a reappraisal of our orientation and of our research efforts in view of the modern advances which surround us.

Overemphasis on Empiricism

It has been the thesis of this text that probably no obstacle stands so clearly in the path of the progress of the science of education, at this stage of development, as does our failure to integrate the multitude of empirical findings which the reams of research studies have produced into meaningful structure. Much of our research efforts thus far have been oriented to what Buswell[16] calls "tinkering," and, though this may be better than empty theorizing, we need to develop both the empirical and the theoretical aspects of research to the point where they give each other mutual support. Our lack of theoretical development has not only prevented us from gaining adequate perspective into various educational problems—which only a theoretical conception of the field can provide—but it has also allowed our research efforts to be exerted in diverse and confusing directions. As a result, our research has been isolated, repetitious, and haphazard rather than continuous and systematic.

[16] Guy T. Buswell, "The Structure of Educational Research," *Phi Delta Kappan*, 24 (December 1941) : 167–9.

Thus far, we have relied too heavily on psychology for the development of theoretical perspective. Psychology is not only acutely aware of the importance of theory in guiding its efforts at the empirical level, but it has also maintained a rather nice balance between theoretical and empirical advances. Considerable progress has been made in the theoretical structure of the psychology of learning; Hull's theory,[17] for example, is particularly ambitious. More recently Guilford's framework of mental and of psychomotor abilities[18,19] is both comprehensive and insightful. In more restricted areas, Dinsmoor postulates an avoidance of punishment theory of behavior,[20] while Restle presents a theory of discrimination learning from which he derives three empirical laws permitting the prediction of the behavior of both rat and human subjects.[21] Psychology still has many conflicts to resolve, but even these conflicts are orienting research in directions that are meaningful and fruitful from the standpoint of its growth as a science.

In contrast, educators have been far too much oriented toward practicality: "If it works, it works; why bother finding out why?; Why seek other methods when, after a lot of work, we might find nothing better? Anyway, education is too complex; it is affected by too many factors to permit generalization." We tend to glorify empiricism as synonymous with science. There is a tendency to consider facts the ultimate form of knowledge—a fact is a fact! A theory, on the other hand, is impractical and essentially equivalent to speculation and guesswork. Schoolmen frequently use the cliche: "This is all right in theory but it won't work in practice," forgetting that to be a good theory it must work in practice or it isn't worthy of being called a theory. Yet they themselves operate on the basis of their own unproven theories, untested and untestable assumptions, dogmatic assertions, and unwarranted generaliza-

[17] Clark L. Hull, *Mathematico-Deductive Theory of Rote Learning: A Theory in Scientific Methodology* (New Haven: Yale University Press, 1940).

[18] J. Paul Guilford, "The Structure of the Mind," *Psychological Bulletin*, 53 (July 1956): 267–93.

[19] J. Paul Guilford, "A System of the Psychomotor Abilities," *American Journal of Psychology*, 71 (March, 1958): 164–74.

[20] James A. Dinsmoor, "Punishment: II. An Interpretation of Empirical Findings," *Psychological Review*, 62 (March 1955): 96–105.

[21] Frank Restle, "A Theory of Discrimination Learning," *Psychological Review*, 62 (January 1955): 11–9.

tions of subjective impressions from their personal teaching experience. They are convinced that good teaching is an art derived from experience and requiring no scientific verification, that a scientific foundation is not necessary or even useful for effective teaching, and that questions of teaching can be resolved by proclamation.

Inadequate Research Methodology

Although a case has been presented for the need for theoretical development, it is equally true that a theory cannot be any better than the empirical findings on which it is based. Unfortunately we are still using research methods that are inadequate for the solution of the problems we face. We frequently act as if we considered a survey of group opinion an adequate approach to educational truth, and even when we experiment, we generally subscribe to the monistic concept of science, though it is essentially inadequate for dealing with the complex variables with which education must cope. We need more rigorous classical designs of research and more adequate analysis and interpretation of results. We are still hampered by our relative unfamiliarity with research and statistical methods of the complexity required for the adequate attack on our problems.

Much of the research conducted in education is faulty. Many studies contain flaws that automatically make them null and void from the standpoint of scientific truth, and potentially dangerous from the standpoint of application. These errors cover all aspects of research: improper formulation of the problem, inadequacy of control, non-representativeness of the sample, invalidity of the data, invalidity of the criterion, inadequacy in the analysis, errors in interpretation, and so on. Marquis lists six types of research occurring, "frequently enough so that they are easily recognizable," that make relatively little contribution to the advancement of science: 1. *wisdom research* which makes a thorough review of the literature but does not get to the point of testing anything; 2. *unfocused research* which goes in all directions at once with no problem to guide it; 3. *practical research* which solves the problem at the local level, but does not contribute anything to the theory of research or the solution of further problems; 4. *de-*

scriptive research, which simply describes a certain phenomenon—for example, opinion polls; 5. *theoretical research,* which, though essential to the development of science, does not suggest ways in which a theory might be tested; and 6. *critical ratio research,* in which statistical manipulations of a correct nature are made, but in which there is a lack of theoretical framework.[22]

Many studies are inconclusive and in conflict with other investigations. Many have been conducted loosely and, it seems, for the purpose of confirming the investigator's previous viewpoint. For instance, Conger notes that, in the field of driver education, poor research and worse reporting have produced some strange results.[23] He suggests that perhaps the only valid conclusion that can be reached in driver-education research is that factors other than driver education itself are the significant contributors to the reported differences in accident and violation rates between driver-education and non-driver-education groups.

Much of the current research has been oriented toward the solution of immediate problems while what is needed is a fresh coherent view stated in a way that permits scientific study, designed for a long-range attack on significant problems. As Fattu points out, there are too many inadequate studies whose only justification is that it was the best that could be done under the circumstances.[24] Of course, inadequacy in research is not the monopoly of education. Every profession tends to be overcritical of its accomplishments. Similar criticism was leveled at psychologists by Ellis,[25] whose survey of psychological research revealed essentially the same inadequacies that are noted in education. He concluded that there was evidence to support the contention of critics that psychological research could well afford a more intensive, co-operative, and fruitful type of planning and execution.

[22] Donald G. Marquis, "Research Planning at the Frontiers of Science," *American Psychologist,* 3 (October 1948) : 430–8.

[23] John J. Conger, "Personality Factors in Driver Education," *Phi Delta Kappan,* 41 (June 1960) : 396–7.

[24] Nicholas A. Fattu in Frank W. Banghart (ed.), *First Annual Symposium in Educational Research* (Bloomington: Phi Delta Kappa, 1960).

[25] Albert Ellis, "What Kind of Research Are American Psychologists Doing?" *American Psychologist,* 4 (November, 1949) : 490–4.

Lag in Educational Practice

A story is told of an agricultural agent who, visiting a farmer in the hinterland and noticing the dilapidated conditions of the farm, told the farmer he would send him pamphlets suggesting improvements he might make. Returning some months later, the agent found the farm in almost exactly the same conditions of disrepair and asked the farmer: "Didn't you get my pamphlets?" Whereupon the farmer replied: "Oh, yeah, I did; but I ain't read them. Why, shucks, I ain't farming half as good as I already know how!"

Probably nowhere does the moral of this story apply better than it does in connection with the dual lag which exists between educational research as we know it should be conducted, as it is conducted, and as it is applied in the classroom.

The limitations of current educational research have been noted. Despite the fact that we know what constitutes adequate research on educational problems, much of the research that is done is based on inadequate research procedures and, thus, yields only partial and inconclusive answers. Experts have devised complex multivariate research designs. We know that these are the only designs adequate for dealing with many of the complex variables found in education; yet only a minority of educators are able to understand such techniques; an even smaller minority are able to use them in their own investigations. Undoubtedly much of the cluttering of the educational journals with trivial research stems from the inability of the writers to conduct the more sophisticated research called for by a more significant topic. A number of writers[26,27] have listed criteria for the evaluation of research; yet, invariably, educational research is criticized for its inadequacies.

An even greater gap exists between research results and their application in the classroom. Despite the hundreds of studies conducted in education and related fields every year, educational practice is frequently based as much on tradition, common sense, and consensus, as it is on research. The reasons

[26] James H. Fox, "Criteria of Good Research," *Phi Delta Kappan*, 39 (March 1958): 284–6.

[27] Dael L. Wolfle, "Standards for Appraising Psychological Research," *American Psychologist*, 4 (August 1949): 320–8.

for this gap are obviously multiple rather than single, but certainly among the more fundamental is the lack of appreciation of the crucial role of research to the advancement of educational practice. The lack of orientation toward research which characterizes educators as a group, is probably even more characteristic of the practitioner who, because of his limited training in research and statistical methods, is frequently cut off from research as an ally in the solution of his problems and as a foundation of good teaching.

The difficulty probably stems in large measure from the lack of orientation toward research of the whole teacher-education sequence, a program in which answers—even at the graduate level—are more frequently given than found, and in which the answers that are given have unwarranted finality and universality of application. As Coombs suggests, we do not change simply because we are not trained to change.[28]

Undoubtedly the lag in educational practice also is related to the inconclusiveness of current educational research findings. Not only are there many educational problems for which we do not have a solution, but even where solutions are available, they are frequently only partial answers whose validity in a given situation is open to serious question. It is understandable that a busy administrator cannot keep up with research on all aspects of educational practice to the point where he can balance one study against another to gain perspective of the validity and applicability of the literature he reads. His attempts at consulting the research literature frequently result in error or frustration, or both. Pertinent sources are often difficult, if not impossible, to find. He is generally pressed for time and cannot chase from one journal to another to find the most adequate study. As a result, he frequently latches on to the first study he finds even though it may not be scientifically sound, typical of the research in the area, or applicable to his present situation. If he presses the subject further, he is likely to discover that White found this, that Black found almost the opposite, and that Gray suggested: "It depends." In the meantime, the articles are poorly written; some fail to emphasize the special factors that led to the peculiar results obtained and there-

[28] Philip H. Coombs, "Education's Greatest Need: A Vice-President in Charge of Heresy," *Phi Delta Kappan*, 41 (March 1960) : 243–7.

fore are misleading; some are based on careless research and are in various degrees of error; many are of an artificial nature. No one has bothered to synthesize the evidence or to resolve the conflicts and contradictions, so that it can be grasped by a person who is primarily an administrator and not a professional researcher.

Not only are many of the studies in various degrees of inadequacy and/or error, which make their findings relatively inapplicable anywhere, but even when conducted under ideal conditions of control, the results still have to be adapted to the local situation. All of this demands a greater competence in the evaluation and interpretation of research than the average practitioner, with his limited research and statistical training, is likely to possess, and this is very discouraging. When he attempts to conduct his own research studies, the practitioner's efforts generally meet with less than complete success. Limitations in time, in facilities, and in research competence on his part and that of participating teachers frequently take their toll, and the outcomes are often disappointing.

Whether as a consequence of such discouragement or as a compensation for their insecurity with respect to research procedures, many teachers and administrators subscribe to "practicality." They cite community opposition to research and the need to "let sleeping dogs lie" and insist, "I don't have to do research to know that . . ." as they continue to use procedures which they have found "effective" in the past. It has also been suggested that educators do not have to meet competition in the way that industry does and that, as a result, they can afford to go on with "half good" procedures. Removed one step further is the classroom teacher, who usually has had very little introduction to educational research in his undergraduate preparation, and who loses contact with research in his everyday practice. As a result, he teaches in much the same way as he was taught, without attempting to relate his practice to the newer advances or even to adapt his procedures to findings of long standing.

The obvious question here is: If educators are opposed to research, specifically how do they propose to solve their problems? The answer seems to be in their reliance on personal experience and, of course, on crudely conducted, control-less

research. This is coupled with "expert" opinions, and, as Ausubel points out, it is generally believed that anyone who has been in the profession for twenty-five years is entitled to some opinion; indeed, he is entitled to make dogmatic pronouncements on pedagogical procedures which require no verification whatsoever and are valid by fiat alone.[29] Many educational practices simply duplicate what some successful teacher has found effective in his particular case—which is assumed to be universal—even though the teacher has frequently only limited understanding of why his methods work.

And yet, educators need to realize that they cannot continue to make the same needless mistakes. Furthermore, if science is not going to save education, what alternative is there? Good intentions are not a substitute for good techniques in achieving desirable goals, for, as Lundberg points out, it is one thing to have a heart in the right place but good intentions must be made operative through effective scientific techniques.[30] It also is time for us to resolve the question of the cost of educational research. And perhaps, we should take our cue from industry, where the cost of research is both large and unquestioned. Perhaps rather than asking "What is the cost of research?" we should ask: "What is the cost of not doing research?" This cost can too often be measured in retardation, in ineffective education, in dropouts and in juvenile delinquency. If we are to continue to invest in education as we have in the past, perhaps it would be wise to ensure the maximum effectiveness of the operation. There is undoubtedly merit in Anderson's suggestion that merely urging teachers to put more effort into their teaching may not be the solution.[31] Rather than beating dead horses, teachers might do better to reallocate some of their efforts to the discovery of more adequate methods.

THE GRADUATE SCHOOL

The graduate school is probably the key to the future of education, for, in no small measure, it determines the kind of

[29] David P. Ausubel, "The Nature of Educational Research," *Educational Theory*, 3 (October 1953) : 314–20.

[30] George A. Lundberg, *Can Science Save Us?* (New York: Longmans, Green, 1946), p. 104.

[31] Kenneth E. Anderson, "Improving the Competence of Teachers in Measurement and Evaluation," *Science Education*, 45 (December 1961) : 424–9.

leadership education will get both at the local and at the national level. The history of the graduate school goes back to Europe—to England and, particularly, to Germany. In America, in the one hundred years since Yale conferred its first Ph.D. degree (1861), we have seen a rapid increase in the need and demand for advanced education until, today, some 300,000 students are enrolled in the various graduate departments of American colleges and universities. (And, this figure is expected to triple in the next twenty-five years.) Of the over 75,000 master's degrees awarded in 1961–1962, nearly half went to students in education. The corresponding figure at the doctoral level is 10,000 degrees, and approximately one-sixth of these were in education.

If it is to discharge its crucial function, the graduate school must define clearly its purposes and its mode of operation. First, though graduate training is superimposed on undergraduate training, it must not be simply a continuation of undergraduate work but must involve a considerable departure both in the degree and in the kind of training it emphasizes. To the extent that graduate work is designed to provide insights and practice in leadership for the profession, graduate studies should not be a matter of regimentation of students to a curriculum of standard courses tailored to undergraduate specifications. Rather than courses geared to the absorption of knowledge, the graduate program must place its emphasis on the development of a person capable of discovering his own answers as the basis for making his own decisions.

The Research Requirement

The graduate school's responsibility for providing graduate students with the research skills necessary to the advancement of education as a science is of particular interest for this text. Almost invariably, graduate programs in education require an introductory course in educational research at the master's level and, frequently, an advanced course at the doctoral level. Such courses generally enroll students who range from those with a good mathematical and scientific background, interested in becoming professional research workers, to those of less adequate background, who simply want to con-

tinue classroom work with increased insight and efficiency. In organizing such courses a distinction is generally made between the consumer and the producer points of view, with an emphasis on the former. Among the topics generally included in such courses are the nature and role of research, the scientific method, the different research methods and research designs, library skills, the collection, analysis and interpretation of data, sampling, statistical inference, the preparation of a research report, and significant research studies.

Obviously, a research course at the introductory level cannot provide an adequate coverage of all the essential research techniques. All it can do is provide an orientation to research methodology and an appreciation of its importance. It then becomes the responsibility of the courses in the student's specialization to pursue the topic further from the standpoint of the applications of research to his special field, and to education in general. In other words, it is not the function of the course in introductory educational research to turn out qualified research workers, but simply to provide a basis on which graduate students can deal with educational problems with a greater degree of scientific insight.

The relative incompetence of educators in research, despite the almost universal requirement of a course in educational research at the graduate level suggests a need for a reconsideration of the purpose, orientation, and content of such a course. It is important to realize, for example, that, if education is to prosper, training in research methods must not be so narrow that it trains research technicians who may know how to apply techniques and procedures, but who do not understand the overall framework within which they are operating. Statistical competence and research methodology are simply tools—the means to an end—contributing to knowledge, not ends in themselves. There is a special need for emphasis on the overall conceptualization of science as the framework within which research operates. Ryans,[32] for instance, recommends, as background for doing work in graduate education, an undergraduate major in sociology or psychology, with a minor in the physical or biological sciences. He would also include, among

[32] David G. Ryans, "Preparation of Educational Research Workers," *Journal of Educational Research,* 49 (November 1955): 195–202.

other requirements, educational philosophy, the history of education, experimental psychology, experimental educational psychology, statistics, educational and psychological measurements, test and questionnaire construction, and the fundamentals of research methods. He strongly recommends experience in the conduct of research, such as might be obtained in a bureau of educational research or of field services, including acquaintance with the use of modern data-processing equipment. He also recommends teaching or related educational experience, and goes so far as to suggest a special degree in research methods with special certification.

Graduate students in education are apparently lacking in research skills. They are also relatively inept at professional writing. These criticisms raise an interesting question: "If our students are inadequate in the various aspects of research, where should they have gotten this background?" A degree can call for only so many credits, are we suggesting additional credits? Inasmuch as other fields—psychology, for example—are apparently able to produce students of reasonable competence in the same period, is the field of education too broad for anyone to be able to develop competence in any one area? There are those who feel that perhaps we are at fault when we require practical experience in the classroom as a prerequisite to graduate work in education. It has been said, for instance, that experience as a classroom teacher is incompatible with the creative and inquisitive mind required of a research worker in that it promotes the development of educational practitioners rather than of investigators.[33] Particularly dangerous from this point of view is the part-time, on-the-job approach to the degree; it is difficult for the student to disengage himself from the demands of the classroom, and the result is often a low level of scholarship.

A related problem is the type of student which graduate work in education attracts. The pressures placed on teachers to improve their qualifications lead many to apply for advanced work, despite their limited suitability. There is a need for careful screening of candidates, particularly at the doctoral level, and, in this selection, it may be desirable to place considerable emphasis on the creative and imaginative mind.

[33] Robert M. W. Travers, *op. cit.*, p. 339.

There is also a need to place greater emphasis on research and research methods, both in the undergraduate and in the graduate programs.

The Thesis Requirement

The trend towards abandoning the thesis as a requirement for the master's degree in education, and a number of the social sciences, is lamented by many educators who look on the thesis as the crowning glory of graduate studies. The reasons for dropping the thesis requirement for the M.Ed. degree are, of course, many and varied. It is argued that the large number of master's students in education make it relatively impossible to provide adequate supervision. It is also claimed that students at the master's level are not sufficiently advanced for them to select a suitable topic within the narrow range of their research and statistical competence, and that, consequently, insisting on a thesis from every master's student produces nothing more than second-rate term papers, or simple clerical reports of frequencies relative to a trivial problem. To the extent that these allegations are true, additional course work and a comprehensive examination might be a superior alternative.

Considered from a logical point of view, however, abandoning the thesis requirement—and accepting a project rather than a dissertation at the doctoral level—implies that research experience is not important in the training of a graduate student in education, that education should be oriented toward the practical rather than toward the scientific, and that even the leaders of the teaching profession are practitioners who need less emphasis on the development of research skills than is required in the traditional M.A., the M.Ed. or Ed.D. with thesis, or the Ph.D. In practice, it frequently means that the distinction between the undergraduate and advanced degrees is simply one of the number of courses and examinations that the student has completed satisfactorily.

The Language Requirement

Another of the traditional requirements for an advanced degree which is getting progressively less emphasis is profi-

ciency in a foreign language. Although there are differences of opinion, the trend toward minimizing the need for a foreign language is, perhaps, more defensible than that concerning the thesis. Inasmuch as nearly all of the worthwhile material concerning education and related fields published in foreign languages is readily available in translated form (in fact, in many libraries it is available only in translated form), it would seem wiser to allow the student to substitute another research tool, such as advanced statistics or advanced research designs. This argument does not apply to certain other disciplines, nor does it apply to the education student working for a degree in comparative education where the use of a foreign language would be a valuable tool. The graduate requirement of a foreign language is sometimes justified on the grounds that, with the present emphasis on the exchange of American and foreign scholars, proficiency in a foreign language is more important than ever. Though effective consultation does depend on facility in communication, it is highly unrealistic to expect a student to have proficiency in all the languages where such exchanges might conceivably take place. It must also be noted that the nations in greatest need of consultants are generally the very nations that have least to contribute in their native language.

Participation in Research

In the physical and biological sciences, graduate students generally work with the faculty members as assistants on various projects. Graduate students in government and economics also frequently engage in community research. In contrast, and though there are notable exceptions in the bureaus of research and of field services of some of the better schools, graduate students and faculty members in education are rarely engaged in research, despite the fact that there are school systems needing help with problems of all degrees of complexity and scope. Presumably research grants are available in education, and research experience would vitalize graduate work. The lack of this kind of work is apparently the fault of the education faculty who too frequently get bogged down in teaching, while the graduate student gets oriented towards research as something to be talked about—once his project is over.

ACTION RESEARCH

As we have seen, teachers are constantly faced with problems which may be best attacked through fundamental research designed to develop the general principles within which such problems exist. Once the appropriate principles have been derived, the solution to the specific problems can be deduced. Although the derivation of the required principles may call for extended and complicated research, once the basic principles are derived, they are applicable to a wide range of sub-problems which no longer need to be investigated individually. Many of the problems facing the educator require immediate attention, however, and frequently it is more expeditious to attack these problems directly rather than by the derivation of broad basic principles. Such on-the-spot research, aimed at the solution of an immediate problem, is generally known in education as *action research.*

In contrast to pure research, which is concerned with the derivation of generalizations of broad applicability and only secondarily in any practical value they might have, action research is undertaken to act as a guide to action in a specific problem area. It is oriented toward the solution of an immediate problem, and it is only secondarily interested in making a contribution toward the discovery of broad generalizations and the development of the theoretical structure within which the problem exists. The action researcher is a practical man who is willing to forego scientific rigor in order to obtain a usable answer to a problem existing at a local level; he is concerned with the situation as it is today on the assumption that this is the kind of situation he will continue to have and that his findings are useful to the extent that he bases them on actual cases.

The person most responsible for the development of the concept of action research is Stephen M. Corey, whose *Action Research to Improve School Practice,*[34] published in 1953, hit a responsive chord among teachers plagued with problems and unfamiliar with the means for their solution. Action research

[34] Stephen M. Corey, *Action Research to Improve School Practice* (New York: Teachers College, Columbia University, 1953).

represents the implementation of Dewey's idea of harnessing classroom teachers in the solution of their particular problems.

Advantages of Action Research

Among the many advantages that can be claimed for action research are the following:

1. The appeal of action research is based, in part, on the relative inability of the pure researcher to communicate with the practitioner. The fundamental researcher apparently cannot be bothered with the mundane problems of the man in the field, or he insists (and correctly so) on developing research designs, with all the statistical trimmings, which are so complex that the practitioner cannot understand them. He appears to be under the impression that if he carries out research and publishes his results, people in the field will be eager and able to put the findings into practice. Frequently, as pointed out by Corey, he prides himself on being a scientist and considers not having to deal with the practical situation a matter of virtue. He occasionally laments the fact that his findings are not being used, and he spends much time discussing the lag between research and practice. On the other hand, the practitioner, often with limited research background, finds it difficult to translate the outcomes of pure research derived under conditions of strict control into an actual classroom program. Action research, in contrast, provides him with solutions he can more readily understand and adopt.

2. The most obvious advantage of action research stems from the fact that any change in teacher behavior and teaching practice must be preceded by a corresponding change in the thinking and in the attitudes of the teacher. Such a change is more likely to take place as a result of research which the teacher helped plan, conduct, and evaluate, than it is as a result of research reported in some journal. Action research is, therefore, more conducive to the implementation of research findings, since it is frequently easier to inaugurate innovations on an overall school basis than to convert each teacher individually. Under proper leadership, co-operative research of this kind utilizes all the advantages of group dynamics in drawing out the best participation of the teachers involved, in overcoming

inertia and defeatism, and in making teachers feel secure while investigating sources of difficulty. The feeling of accomplishment at having tackled their problems constructively—at their own level—often is a morale booster and a revitalizer of teachers, frustrated at having to face problems and not being able to do anything about them.

Teachers gain through seeing that others have problems similar to theirs; they see that having problems is not a sign of incompetence, that problems are not something one excuses and denies but something that one solves. Combining their talents for the solution of mutual problems frequently results in a feeling of partnership in scholarship, an improved group feeling and enthusiasm, and a higher level of research consciousness. Action research frequently leads to the solution of problems that could not be solved without the complete and wholehearted participation of the whole faculty.

3. Discussions connected with the planning stages of action research are generally very helpful in providing teachers with insight into the nature of educational problems and of research techniques and in stimulating them to read the related professional literature. The general familiarity with their immediate problem which they gain is likely to develop in teachers a greater understanding of the problems of the classroom and a greater competence in deriving solutions, both from the published literature and from more adequate investigations of their own.

Limitations of Action Research

Action research is subject to a number of limitations and pitfalls which must be recognized. Since action research is almost completely empirical and local in nature, its contributions to the development of education as a science are likely to be secondary. Under optimum conditions, it can contribute facts to be integrated into theory, it can provide for the testing of theory and the possibile verification and refutation of theoretical concepts, it can result in the clarification of theory, and it may eventually help integrate previously existing theories. Unfortunately, however, the maximum benefits from action research are seldom realized because of a failure to generalize the results and to integrate them with the previously existing

theoretical structure. A number of other limitations are more directly connected with the method itself.

1. A major limitation of action research is its relatively poor quality. Teachers generally are not researchers, and they are likely to experience a number of difficulties in obtaining meaningful results, especially since action research involves a maximum of flexibility in the operation and interaction of the multiple variables in the situation. The difficulties may arise from failure to define the problem clearly, resulting in the gathering of tons of data without the guidance of a hypothesis; inability to control extraneous factors; inadequacies in the treatment of the data; misapplication of the results; and so on. Often it is difficult to attain academic usefulness while maintaining scientific adequacy. Frequently, teachers undertake over-ambitious projects and expect results too soon, and, as a result, become disappointed and discouraged. Because of the failure to maintain adequate control, it is often difficult to identify causes, even when results are obtained. Action research can become a case of the blind leading the blind, and the problem is further aggravated by that the fact that teachers generally are too close to their problems and too untrained in scientific objectivity to be rigorous and objective in their approach.

The possible sources of difficulty point to the need for a consultant working closely, both directly and through a steering committee, with the teachers in order to promote greater compliance with recognized principles of scientific research. This consultant needs to be highly trained in public relations and group dynamics, as well as in research methods, and must provide close supervision if he is to keep untrained and ego-involved teachers on the track.

2. A major consideration in action research concerns the generalizability of its results. Fundamental research starts by defining a population and, then, takes a representative sample to serve as the basis for inferences concerning this population. In contrast, action research starts with a sample the nature of which is not identified but simply taken as is. It is not clear, therefore, to what population the conclusions and insights reached in the study are to apply. For example, there is an apparent assumption that the teacher's present class is sufficiently representative of his future classes that the present results can be ap-

plied legitimately to the groups he is likely to have in the years to come. As pointed out in Chapter 7 this sort of *populationing* is always risky, especially since frequently in action research the problem is ill-defined, the procedures used are left unclear, the subjects unidentified, and so on. The applicability of the findings to another school, in the event of teacher transfer, is even more questionable. In fact, action research frequently violates the basic rules of scientific research; not only is it generally conducted in an atmosphere of common sense rather than of scientific control, but it is frequently seriously lacking in the extent to which the various criteria of rigorous research are met.

3. Another criticism of action research is that it frequently is added to the shoulders of already busy teachers who have only limited freedom to say no. The result may be inadequacy both in their teaching and in their research. This is not an insurmountable problem: if we believe in the value of this type of research, ways can be found to release teachers from part of their other responsibilities. It is also possible to have teachers work on problems which are close enough to the problems which they face so that they would want to do this research as part of their regular responsibilities. Much depends on the leadership provided by the principal; teachers are generally willing to do research if they see that it will help them meet their problems. It would seem crucial, however, that teachers not feel they *must* do research.

Despite its limitations, action research is certainly to be encouraged. Participation by teachers in the solution of their problems is to be encouraged. Action research has led to the solution of many classroom problems, and it has contributed to the advancement of education as a science by providing tentative hypotheses and tentative generalization of immediate practicality. Action research also raises the professional caliber of the participants. Furthermore, even though some action research is probably not research at all (any more than the cooking of an egg deserves to be called culinary research), but rather a type of professional activity a professional person would normally do in the dispatch of his responsibilities, one must be careful not to make a dichotomy of action and fundamental research. The distinction—whether considered on the

basis of the method or quality of the research or on the breadth and immediacy of the applicability of the results—simply has reference to the two ends of a continuum. The difference is essentially one of emphasis. On the other hand, even when it is conducted under the close supervision of a competent consultant, the contribution of action research to the promotion of education as a science is likely to be of a relatively secondary nature. There is need, therefore, while encouraging action research, to recognize the importance of parallel fundamental research.

RESEARCH BUREAUS

Probably nothing more eloquently spells out the lack of orientation of school administrators toward research than does the low status of the research bureaus which they maintain. School systems, with financial disbursements in the millions, almost invariably are content with staffing their "research departments" with one or two clerks who, despite such titles as "Research Director," do the bulk of their "research" by tabulating attendance records and attending to other clerical chores.

It is generally recognized that whatever else can be said for or against research bureaus as they are operated by many school districts, they cannot be accused of conducting research. The few studies that have been made of their functions and operations point to the fact that "research bureaus are not doing research." Most of them would be better labeled pupil accounting or testing bureaus. Although there are a few outstanding exceptions, their responsibilities are generally more clerical than professional. Even state departments of education and state educational associations, from which real leadership might legitimately be expected—and let us not minimize the usefulness of the data which they collect—are not doing the type of basic research that needs to be done if schools are to fulfill their obligation to society.

Role of the Research Bureau

We have seen that if educators do not do research, no one is going to do it for them. We also know that teachers do not have the time or the know-how to conduct the type of complex research that needs to be done if education is to solve its prob-

lems. A major solution, it would seem, lies in organizing within each school system a research bureau capable of providing aggressive leadership in the solution of educational problems. Since many of the problems—practical or theoretical—faced by teachers are very frequently those faced by other teachers in the system, only a system-wide attack on such problems will lead to their adequate solution. It seems reasonable, therefore, to expect the central office to initiate and to co-ordinate such research and become the logical center of operations. The problem of integrating bilingual children into the school, for example, can be handled only on a system-wide basis. Many other problems fall into the same category.[35]

Though it is true that classroom teachers generally cannot conduct high-level research on their own, it must be recognized that at no time in our history have teachers been as well qualified and as professionally minded as they are now. If they are capable of teaching children, they are also capable of doing research that can make their teaching more effective. They need to be encouraged to become research conscious and to be guided in their efforts to deal with their problems systematically. Under the proper leadership, they can be involved both in the planning and in the conduct of research, and the implementation of its outcomes. When this is done, research is no longer an additional burden placed on the shoulders of already overworked teachers; it becomes part of their job integrated with the task of teaching. As a result of their involvement in first-class research, teachers generally find that education takes on a new meaning and teaching becomes a true profession.

A central research bureau could assume leadership for a number of research projects to be conducted in any number of individual schools, depending on their nature and their scope. The central bureau might be expected to discourage the choice of relatively impossible topics, to help with the formulation of the research design, to facilitate the organization of control groups, to enlist the co-operation of the necessary personnel, to specify the part each is to play and to assign definite responsibilities, to provide moral support and consultation, and finally to provide for the dissemination of the results and for their implementation. University personnel could be expected

[35] Some of these problems are better tackled at the state level—or perhaps even at the national level under the auspices of the U.S. Office of Education.

to provide consulting services. This would enable the education faculty to keep in contact with the school so that education is no longer something that one teaches, and further would provide graduate students with field experiences. Such a team —research bureau personnel, classroom teachers, university consultants, and graduate assistants—should be an adequate combination for a successful attack both on the immediate and on the long-range problems of the teacher.

Research Bureaus at the University, State, and National Level

So for our setting has been the public schools; the need for research bureaus also applies to the college situation and, even more specifically, to the school of education, where problems frequently are met only at the discussion level, where what little research is conducted is done in a piecemeal fashion, where half-vast problems are tackled with half-adequate techniques by faculty members—on stolen time and minus facilities —as they run between classes and committee meetings. Both Clarke[36] and McArthur[37] found that research on educational problems conducted at the university level is of the same haphazard nature as that in the public schools. Here too, there is a need for an agency— a bureau of institutional or educational research—to co-ordinate the research on the thousand and one local problems, ranging from the admission of students, to the follow-up of former students. Such a bureau can also engage in theoretical research that will raise education from the empirical to the more sophisticated levels of science, though theoretical research should probably be subsidized by outside grants rather than by the local university funds. When staffed by adequate personnel, such a bureau can provide real leadership in the promotion of education as a science. It can draw on all university personnel and can make all of its facilities available to the schools and other agencies. It can serve as a training ground for graduate students and, thus, contribute to the training of future educators capable of assuming their places among their fellow scientists. Probably no finer tool for upgrading the profession can exist than such a bureau operating

[36] Stanley C. T. Clarke, "Trends and Problems in Educational Research," *Alberta Journal of Educational Research*, 3 (December 1957): 209–19.

[37] R. S. McArthur, "Organization for Educational Research in Universities of Midwestern United States," *Alberta Journal of Educational Research*, 4 (September 1958): 131–41.

effectively in the solution of educational problems and in the development of the future leaders of the profession. It can also provide consulting services to the various departments of the university, facilitate communication among research workers, provide space and facilities for research activities, and generally help to co-ordinate the research activities of the university.

A similar need for research leadership exists at the state and national levels. Administering education is big business, involving an annual expenditure of millions of dollars. Industry spending that kind of money would want to make sure it is spent in the most efficient way possible and would allocate a considerable percentage of its outlay to research. At the state level there is need for a bureau of educational research to provide leadership for the research activities of the state department of education and the school systems under its jurisdiction, to co-ordinate research efforts within the state and from one state department of education to another, and to promote general improvement in educational practice through the dissemination and implementation of research findings.

The United States Office of Education can be considered a super-bureau of educational research—leading and co-ordinating the research efforts of the nation—and it might be expected to continue and to accelerate its efforts in these directions. Educational problems are frequently nation-wide in nature and, consequently, require a nation-wide attack. It might avail itself, to a greater extent, of the services of professional and lay organizations interested in the advancement of education for defining and structuring significant areas in need of a long-range attack, which generally are beyond the resources of local, and even state, organizations. It needs to continue to expand its present Cooperative Research Program of financial assistance to worthy educational research projects.[38]

REORGANIZATION OF OUR RESEARCH EFFORTS

Need for Reorganization

Education has made only limited progress in the resolution of its numerous and complicated problems. While this in itself

[38] United States Office of Education, *Cooperative Research Projects.* (Washington, D.C.: The Office of Education, Yearly, 1957–date.)

is no cause for alarm, the analysis presented in this chapter of the current status of education, and of research as the means by which education is to be furthered as a science, suggests a definite need for a reorganization of our research efforts. If education is to participate in the scientific advances that characterize the age in which we live, we can no longer continue to place major reliance for educational research on the haphazard and incidental efforts of the amateur, the hobbyist, and the graduate student. Such an approach is not—and cannot be—adequate for the demanding task of ensuring our scientific progress. Rather, we need to devote ourselves as a profession more energetically to a planned and systematic attack on our problems, for as Fattu emphasizes, "only by inspired, sustained, and systematic research in education similar to that which has graced the other sciences can education become truly effective."[39]

In contrast with the other major disciplines which have placed research in the hands of the professional researcher, we are sporting our freshman team in a field so vital to our progress. Furthermore, we are providing graduate students—on whom we have depended for a large portion of the research conducted in education—with only limited coaching from relatively inexperienced coaches, who divide their time between teaching, advising students, writing books and articles, and attending meetings and conventions. To make matters worse, we frequently promote the coaches to administrative positions so that, as soon as they develop proficiency in research, their research activities are curtailed. There is need for a reconsideration of our present over-reliance on the doctoral student as the standard bearer of education in its attempt at scientific growth.

There is also a need to encourage the faculty of education to take a more active part in worthwhile research projects through providing for their partial relief from teaching responsibilities. While the faculty of the physical sciences is expected to share its efforts and talents between research and teaching, the faculties of the colleges of education have failed to give recognition to the complementary role of research and

[39] Nicholas A. Fattu, in Raymond O. Collier and Stanley M. Elam (eds.), *Research Design and Analysis*. Second Symposium on Educational Research (Bloomington: Phi Delta Kappa, 1961).

good teaching. There is also need to encourage more active participation on the part of teachers in the field in the solution of the problems they face from day to day.

Emphasis on Systematic and Continuous Research

All these measures are at best only partial, stop-gap, and essentially inadequate approaches to the solution of educational problems, for they are based on the incidental, haphazard efforts of the amateur and cannot be depended on to provide the basis for systematic, continuous, and vigorous educational progress. A particularly strong statement against deluding ourselves that adequate educational progress can come from our present "shoestring" operation was presented by the research committee of the American Educational Research Association in the following recommendation:

> Promote the notion that research is difficult and is best done by the professional. Educational research can probably be promoted best if it is advertised as being hard, demanding, consuming, and requiring a lengthy period of preliminary professional training. Such a perception of research is standard when one thinks of physics, medicine, chemistry, but not so when education is considered. The popular interpretation of action research has so deluded public school teachers, supervisors, and administrators that they believe, first, that the required abilities are destroyed automatically when one announces his intention to do research, and, second, that correlating the distance bus students travel to school and their IQ's epitomizes educational research in its most complex and penetrating aspects. These fallacies are promoted by some research textbooks by assertions that classroom teachers should conduct research (as if teaching were not a full-time job in itself), and by irresponsibility on the part of administrators and supervisors who, when confronted with the need for in-service training of teachers, call it "research" to make it more acceptable to the teachers.[40]

Adequate educational progress can be promoted only through the implementation of long-range, comprehensive, and co-ordinated research programs as exemplified by the investigations of Gesell, Terman, Thurstone, Cattell, Guilford,

[40] Alonzo G. Grace, *Recommendations of the Committee on Research.* Annual Report, American Educational Research Association. (Washington, D.C.: The Association, 1962), p. 26.

and Barr, the Eight-Year Study of the Progressive Education Association, The Ryans Teacher Characteristics Study, and others, reviewed in Chapter 15, or mentioned in various sections of this text. Eurich, for example, recommends setting up a limited number of major educational research and development centers patterned after those in physics, medicine, agriculture, and child growth and development to spearhead the movement for the advancement of education.[41] In addition to being staffed by the nation's top researchers in the relevant academic disciplines working as a team, these centers could also draw on the full research and financial resources of the nation.

At the local level, smaller research units could be formed, each dealing with areas of special interest. These too would have to be staffed by relatively competent researchers devoting a good part, if not all, of their time and energy to the systematic study of educational problems. Grace, for example, points to the growing conviction among research workers that "the minimum allocation of time to a faculty member for productive research is half-time. Lesser amounts of time are apparently too difficult to protect from intrusion and distraction."[42] The successful operation of such a unit would call for the full cooperation of the members of a given department centered around a topic of common interest and continued over a period of years. This would avoid the major objection to short-term discontinuous studies of isolated topics, which is characteristic of our present research efforts.

Structuring of Research Activities

One of the first steps that needs to be taken in the reorganization of our research efforts is the further clarification of the status of the major aspects of education so that we can get a clear perspective of our present position. The amount of material—of varying degrees of quality—that has accumulated helter-skelter in certain areas is so extensive that it is no longer possible to read it all, let alone digest its import. In many cases, it can only be appraised in the words of Blake and Mouton, "Gad, what a mess!"[43] As pointed out by Underwood, unless

[41] Alvin C. Eurich, *op. cit.*

[42] Alonzo G. Grace, *op. cit.*

[43] Robert R. Blake, and Jane S. Mouton, "Personality," *Annual Review of Psychology,* 10 (1959) : 203–32.

the data are synthesized periodically, the field is likely to become progressively more forbidding to anyone interested in it, and consequently it will scare away further research, particularly fundamental research.[44]

Actually, what is needed is not just a summary of the literature but an integration of the vast accumulation of isolated knowledge in the different areas into a theoretical framework, which will place empirical data in a meaningful structure, deepen our understanding of their significance, and permit their more effective use in practice. Particular emphasis must be given to identifying areas and directions in which the research efforts of the profession might be most profitably exerted.

Such a task constitutes a major undertaking, for it calls for particular insight into a given field and special talent at organization—as well as a willingness to wade through loads of material. This might be an area in which some association interested in the advancement of education could do the cause of education a real service. It would help, for instance, if each professional society assumed responsibility for providing yearly abstracts of significant studies in its field, as well as an orientation to its present status, its gains, and its trends, as a means of giving the field perspective and clarifying the nature of the specific issues and problems it presents. It would be of great benefit for education to have a publication comparable to the *Annual Review of Psychology* in which yearly reviews of the major areas would be made. There is also much merit in Conant's suggestion that scholars interested in undertaking a synthesis of the literature in some area be subsidized in the same way as regular investigators.[45] Undoubtedly, such attempts at structuring basic areas of education would be of definite help in guiding the efforts of the profession in its search for more adequate answers to more significant problems and would constitute one of the major contributions that could be made to the cause of education.

If we are to continue placing considerable responsibility for the research that is done on the shoulders of the graduate student and the incidental researcher, we need to channel their

[44] Benton J. Underwood, *Psychological Research* (New York: Appleton-Century-Crofts, 1957), p. 290.

[45] James B. Conant, "The Role of Science in Our Unique Society," *Science*, 107 (January 1948): 77–83.

efforts more definitively into a systematic program, which has significance and continuity. This might best be effected through more active leadership on the part of the profession in defining and structuring areas in need of research at the individual and group level. The leaders of the profession can render a valuable service by orienting research toward crucial areas and by enlisting financial help for comprehensive studies of the scope of the studies previously mentioned. The continuity required for an effective attack on educational problems might also be effected through the co-ordination of a series of studies on the same topic, perhaps involving a number of doctoral candidates working under the supervision of an advisor, or as part of a team in a long-range project sponsored by the faculty of a given school. Neither of these two suggestions needs to curtail the freedom and initiative of anyone who wants to work on a topic of his own; it certainly does not mean that every doctoral student will be relegated to the role of a member of the "machine" taking over where the "just-graduated" Ph.D. or Ed.D. left off. But it would permit greater co-ordination of the profession's efforts, and many research workers (especially the neophytes) would welcome being part of a team working toward a significant objective. It would also provide worthwhile training in teamwork in research. Similarly, action research, useful as it may be at the local level, needs to be made more effective by being structured more definitively through more adequate leadership from the central office and integrated more closely with fundamental research and theory.

If the next war is to be won in the classroom rather than on the battlefield, it behooves American society to provide the means for the improvement of education and educational research, just as it behooves the profession to equip and organize itself for effectiveness and productivity. Funds are apparently available from numerous sources for the purpose of subsidizing worthwhile research. The American public has always been able to support any program in which it believes; it is time that we deserve and get this support.

SUMMARY

1. While education has made substantial progress in recent years, it has not kept pace with the rapid scientific progress which has characterized the twentieth century.

2. Our deficiencies can be attributed to such factors as (1) a lack of orientation toward research as the vehicle for scientific progress; (2) a lack of theoretical framework to structure the empirical findings and to orient the research efforts of the profession; (3) a lack of emphasis on research in the teacher-education program; (4) a lack of familiarity with adequate research methods; (5) an inability to translate research findings into practice; and (6) an overemphasis on empiricism and practicality. It also seems that the pressure of having to solve immediate problems has prevented us from undertaking a more systematic and continuous approach to research.

3. The graduate school holds the key to the future of education. Of particular interest here is the persistent criticism of education's lack of orientation and competence in research despite the almost universal research requirement for the graduate degree. There is a need for a reconsideration of the function of the graduate—as well as the undergraduate—program in education and its relation to research. The de-emphasis of the thesis and the dissertation and the implication that the education of even the leaders of the teaching profession should be practical rather than scientific also bears reconsideration.

4. In contrast to pure research, action research is oriented to the solution of immediate problems and only secondarily to the discovery of generalizations of broad applicability. Action research presents a number of obvious advantages—for example, teacher participation in research is particularly conducive to the implementation of the results and, thus, to the upgrading of educational practice. On the other hand, it has inherent weaknesses that need to be recognized. If they are to be effective in action research, teachers will need considerable guidance from a capable consultant working closely through a steering committee.

5. The relative lack of orientation of educators toward research is particularly evident in the inadequacy of the research bureaus. There is need for a greater leadership to be exerted through genuine research bureaus maintained in the school system, and at the college, state, and national level.

6. If education is to prosper and keep pace with the world its efforts have helped to create, there is need for a reorganization of its research efforts toward (1) more systematic and continuous research conducted under the direction of professional researchers; and (2) a synthesis of empirical findings to clarify the present status of the major problem areas of education, with particular emphasis on the development of theoretical structure and the reorientation of the research efforts of the profession toward meaningful investigation of significant problems. There is a special need for theory-oriented research such as that which characterized early research in the psychology of learning.

PROJECTS and QUESTIONS

1. Make a survey of school bulletins to determine the research requirements for the graduate degree in education at some of the country's leading institutions.
2. On the assumption that the quality of the dissertations it accepts is indicative of the orientation of a given graduate school toward research, through *Dissertation Abstracts* trace the schools with a strong emphasis on research. Verify your judgment through a survey of their research requirements as in (1) above.
3. While some of the big industrial and commercial firms of the turn of the century are no longer, some young companies have grown into modern giants. To what can their success be attributed? Give specific examples by surveying their financial reports for evidence of emphasis on such things as research and development.
4. Check the adequacy of research bureaus maintained by nearby school systems and state departments of education. Specifically, what studies have they conducted? What is their budget and what part of that budget goes for actual research?
5. Identify a problem amenable to action research and plan its execution. Consider carefully the public relations aspects and the effective use of personnel.

SELECTED REFERENCES

ABELSON, HAROLD H. "The Role of Educational Research in a Democracy," *Journal of Educational Sociology,* 21 (April 1948): 454–61.

AHRENS, MAURICE R. "Curriculum Improvement through Action Research," *High School Journal,* 39 (April 1956): 364–9.

ANDERSON, KENNETH E. "Avenues for the Improvement of Research," *Science Education,* 45 (December 1961): 418–24.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. *Symposium on Basic Research.* Washington, D.C.: The Association, 1959.

ASSOCIATION FOR SUPERVISION AND CURRICULUM DEVELOPMENT. *Research for Curriculum Improvement.* 1957 Yearbook. Washington, D.C.: The Association, 1960.

AUSTON, JOHN T. "Rationale for Reorienting Educational Research," *Journal of Educational Research,* 47 (January 1954): 337–49.

BARR, ARVIL S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," *Journal of Experimental Education,* 16 (June 1948): 203–83.

———. "Characteristics of Successful Teachers," *Phi Delta Kappan,* 39 (March 1958): 282–4.

BARR, ARVIL S., *et al.* "Wisconsin Studies of the Measurement and

Prediction of Teacher Effectiveness: A Summary of Investigations," *Journal of Experimental Education,* 30 (September 1961) : 1–56.

BEACH, NORTON L. "Research Goes into Action," *Journal of Educational Research,* 47 (January 1954) : 351–8.

BERELSON, BERNARD. *Graduate Education in the United States.* New York: McGraw-Hill, 1960.

BETTS, EMMETT A. "Phonics: Practical Considerations Based on Research," *Elementary English,* 33 (October 1956) : 357–71.

BETTS, EMMETT A., *et al.* "Unsolved Problems in Reading: A Symposium," *Elementary English,* 31 (October 1954) : 325–38.

BLACKWOOD, PAUL E. and BROWN, KENNETH E. "Science Education Research Studies—1954," *Science Education,* 39 (December 1955) : 372–89.

BLUM, Fred H. "Action Research: A Scientific Approach?" *Philosophy of Science,* 22 (January 1955) : 1–7.

BORG, WALTER R. "Teachers as Intelligent Consumers of Research," *School and Society,* 73 (June 9, 1951) : 357–9.

BRICKMAN, WILLIAM W. "Obstacles Impeding Professional Reading," *Phi Delta Kappan,* 38 (June 1957) : 369–73.

BROWN, KENNETH E., *et al. Analysis of Research in the Teaching of Mathematics.* Periodic Report. Washington, D.C.: United States Office of Education, 1955–date.

BROWNELL, WILLIAM A. "Criteria of Learning in Educational Research," *Journal of Educational Psychology,* 39 (March 1948) : 170–82.

———. "A Critique of Research on Learning and on Instruction in the School," in *Graduate Study in Education.* 50th Yearbook, National Society for the Study of Education, Pt. 1. Chicago: University of Chicago Press, 1951.

BUCK, JACQUELINE V., *et al.* "Review of Recent Research in the Teaching of Science at the Elementary School Level," *Science Education,* 39 (December 1955) : 336–43.

BUCKINGHAM, BURDETTE R. "The Accomplishments and the Promise of Educational Research in Increasing the Over-all Value of Education to the Nation," in *Improving Educational Research.* Washington, D.C.: American Educational Research Association, 1948.

BURMESTER, MARY A. "Behavior Involved in the Critical Aspects of Scientific Thinking," *Science Education,* 36 (December 1952) : 259–63.

BUSH, GEORGE P. and HATTERY, LOWELL H. (eds.). *Teamwork in Research.* Washington, D.C.: American University Press, 1953.

CARMICHAEL, LEONARD, *et al.* "Symposium: Child Development Research—The Next Twenty-Five Years," *Child Development,* 31 (March 1960) : 191–208.

CARMICHAEL, OLIVER C. "The Strategic Role of Graduate Education," *Education Forum,* 26 (November 1961) : 5–14.

———. *Graduate Education.* New York: Harper, 1961.

CARROLL, JOHN B., *et al.* "Neglected Areas in Educational Research," *Phi Delta Kappan,* 42 (May 1961): 339–46.

CARTER, PAUL, *et al.* "The Classroom Teacher and Action Research," *Educational Leadership,* 11 (May 1954): 465–70.

CHASE, FRANCIS S. "The Status of Research in Education," *School Review,* 62 (November 1954): 457–64.

CHEIN, I., *et al.* "The Field of Action Research," *American Psychologist,* 3 (February 1948): 43–50.

CLARK, DAVID L. and CARRIKER, WILLIAM R. "Educational Research and the Cooperative Research Program," *Phi Delta Kappan,* 24 (March 1961): 226–30.

COLADARCI, ARTHUR P. "Are Educational Researchers Prepared to Do Meaningful Research?" *California Journal of Educational Research,* 5 (January 1954): 3–6.

———. "Towards More Rigorous Educational Research," *Harvard Educational Review,* 30 (Winter 1960): 3–11.

COLADARCI, ARTHUR P., *et al.* "Research Priorities in Educational Administration," *Journal of Educational Research,* 47 (April 1954): 625–30.

COOMBS, PHILIP H. "Education's Greatest Need: A Vice-President in Charge of Heresy," *Phi Delta Kappan,* 41 (March 1960): 243–7.

COOPER, DAN H. "The Impact of Research on Education," *Phi Delta Kappan,* 35 (October 1953): 16–20.

COREY, STEPHEN M. "Curriculum Development through Action Research," *Educational Leadership,* 7 (December 1949): 147–53.

———. "Fundamental Research, Action Research, and Educational Practices," in *Growing Points in Educational Research.* Washington: American Educational Research Association, 1949. pp. 261–6. Also *Teachers College Record,* 50 (May 1949): 509–14.

———. "Conditions Conducive to Curricular Experimentation," *Educational Administration and Supervision,* 36 (April 1950): 209–16.

———. "Educational Research and the Solution of Practical Problems," *Educational Leadership,* 9 (May 1952): 479–84.

———. "Action Research by Teachers and the Population Sampling Problem," *Journal of Educational Psychology,* 43 (October 1952): 331–8.

———. *Action Research to Improve School Practice.* New York: Teachers College, Columbia University, 1953.

———. "The Support of Research in Education," *Teachers College Record,* 59 (December 1957): 129–36.

CORNELL, FRANCIS G. "Productive Techniques in Research," *Phi Delta Kappan,* 35 (October 1953): 29–34.

CURLE, ADAM. "A Theoretical Approach to Action Research," *Human Relations,* 2 (July 1949): 269–80.

CURTIS, FRANCIS D. "Milestones of Research in the Teaching of

Science," *Journal of Educational Research,* 44 (November 1950): 161–78.

DOMAS, SIMEON J. and TIEDEMAN, DAVID V. "Teacher Competence: An Annotated Bibliography," *Journal of Experimental Education,* 19 (December 1950): 101–218.

DUKER, SAM. "Educational Research—A Science?" *Phi Delta Kappan,* 37 (April 1956): 318–9.

DURFEE, MACINE and GREENLEE, JULIAN. *Elementary School Science: Research, Theory, and Practice.* Washington, D.C.: Association for Supervision and Curriculum Development, 1957.

ELAM, STANLEY M. "What—and Where is Educational Research?" *Phi Delta Kappan,* 41 (March 1960): 241–2.

EURICH, ALVIN C. "New Dimensions in Educational Research," *A.E.R.A. Newsletter,* 13 (April 1962): 4–10.

EVANS, HUBERT M. and COREY, STEPHEN M. "Cooperative Research and Curriculum Improvement," *Teachers College Record,* 51 (April 1950): 407–74.

FATTU, NICHOLAS A. "Status and Direction of Present Day Educational Research Methodology," *Phi Delta Kappan,* 40 (June 1959): 377.

FOSHAY, ARTHUR W., *et al.* "The Bureau of Educational Research and the Problem of Leadership," *Educational Research Bulletin,* 31 (November 1952): 197–224.

FULKERSON, GLEN. "A Resumé of Current Teacher Personnel Research," *Journal of Educational Research,* 47 (May 1954): 669–81.

GIBB, E. GLENADINE. "A Review of a Decade of Experimental Studies which Compared Methods of Teaching Arithmetic," *Journal of Educational Research,* 46 (April 1953): 603–8.

———. "A Selected Bibliography of Research in the Teaching of Arithmetic," *Arithmetic Teacher,* 1 (April 1954): 20–2.

GOETHALS, GEORGE W. "A Framework for Educational Research," *Harvard Educational Review,* 28 (Winter 1958): 29–43.

GOOD, CARTER V. "Educational Research After Fifty Years," *Phi Delta Kappan,* 37 (January 1956): 145–52.

GRACE, ALONZO G. (Chairman). *Annual Report,* American Educational Research Association. Washington, D.C.: The Association, 1962.

GRAY, WILLIAM S. "Implications of Research for the Improvement of Reading," *Education,* 70 (May 1950): 539–47.

———. "Summary of Reading Investigations," *Journal of Educational Research,* February issue, Yearly, 1924–1960. Continued, Robinson, Helen M. (1961) and Harris, Theodore L. (1962–date).

HANNA, PAUL R. "Proposed: A National Commission for Curriculum Research and Development," *Phi Delta Kappan,* 42 (May 1961): 331–8.

HENRY, DAVID D. "New Priorities in Research," *Educational Record,* 41 (April 1960) : 148–53.

HILGARD, ERNEST R. "Educational Research: Some Friendly Suggestions," *California Journal of Educational Research,* 4 (March 1953) : 51–5.

HODGKINSON, HAROLD L. "Action Research—A Critique," *Journal of Educational Sociology,* 31 (December 1957) : 137–53.

HOLMES, HENRY W., *et al. Educational Research: Its Nature, Essential Conditions, and Controlling Concepts.* Series 1, Vol. 3, No. 10. Washington, D.C.: American Council on Education, 1939.

HUNNICUTT, CLARENCE W. "Team Research in Education," *California Journal of Educational Research,* 4 (May 1953) : 99–102.

HUNNICUTT, CLARENCE W. and IVERSON, WILLIAM J. (eds.) . *Research in the Three R's.* New York: Harper, 1958.

HUNT, HEROLD C. "Educational Research and National Education Policy," *Journal of Educational Research,* 49 (May 1956) : 641–8.

JACOBSON, P. B. "Research Shows Twelve Changes in Fifty Years of Secondary Education," *Nation's Schools,* 49 (January 1952) : 38–41.

JAMES, C. EVAN. "Trends in Child Development Research," *Childhood Education,* 29 (October 1952) : 73–6.

KIDD, CHARLES V. *American Universities and Federal Research.* Cambridge: Harvard University Press, 1959.

KOWITZ, GERALD T. "The Change and Improvement of School Practices," *Phi Delta Kappan,* 42 (February 1961) : 216–9.

LAMKE, TOM A. "Prevalence and Content of Graduate Courses in the Methodology of Educational Research," *Journal of Educational Research,* 49 (October 1955) : 103–11.

LAUWERYS, JOSEPH A. "Definition and Goals of Professional Reading," *Phi Delta Kappan,* 38 (June 1957) : 365–8.

LEVIN, HARRY. "A New Perspective on Teacher Competence Research," *Harvard Educational Review,* 24 (Spring 1954) : 98–105.

MALLINSON, GEORGE G. "The Implications of Recent Research in the Teaching of Science at the Secondary-School Level," *Journal of Educational Research,* 43 (January 1950) : 321–42.

———. "The Five Most Needed Research Investigations in the Teaching of Science and in the Teaching of Mathematics," *School Science and Mathematics,* 54 (June 1954) : 428–30.

———. "The Reading Difficulty of Some Recent Textbooks for Science," *School Science and Mathematics,* 57 (May 1957) : 364–6.

MALLINSON, GEORGE G. and BUCK, JACQUELINE V. "Some Implications and Practical Applications of Recent Research in the Teaching of Science at the Secondary-School Level," *Science Education,* 38 (February 1954) : 58–81.

MALLINSON, GEORGE G., *et al.* "The Reading Difficulty of Textbooks

in Elementary Science," *Elementary School Journal,* 50 (April 1950): 460–3.

HORACE MANN–LINCOLN STUDY GROUP. "Recommended: Group Research for Teachers," *Teachers College Record,* 50 (November 1948): 108–13.

MEAD, ALBERT R., *et al.* "Proposals for the Improvement of Educational Research," *Journal of Educational Research,* 42 (September 1948): 1–7.

MILES, VADEN W. "Bibliography with Annotations of College Science in General Education, 1951–1953," *Science Education,* 38 (December 1954): 366–90.

MONROE, WALTER S. "How Has Educational Research Contributed to the Development of Teacher Education?" *Journal of Teacher Education,* 2 (March 1951): 60–3.

MORSE, ARTHUR D. *Schools of Tomorrow—Today.* Garden City: Doubleday, 1960.

NATIONAL ASSOCIATION FOR RESEARCH IN SCIENCE TEACHING. "Symposium: Recent Research in Science Education," *School Science and Mathematics,* 57 (November 1957): 595–628.

NATIONAL EDUCATION ASSOCIATION, DEPARTMENT OF CLASSROOM TEACHERS AND AMERICAN EDUCATIONAL RESEARCH ASSOCIATION. *What Research Says to the Teacher.* (Series still incomplete). Washington, D.C.: The Association, 1953—

OLGIN, PHILIP. "Let's Re-Examine Progressive Education," *Phi Delta Kappan,* 38 (May 1957): 309–13.

PAGE, ELLIS B. "Educational Research: Replicable or Generalizable, *Phi Delta Kappan,* 39 (March 1958): 302–4.

PHI DELTA KAPPA, COMMISSION ON RESEARCH. "Educational Research: A New View," *Phi Delta Kappan,* 35 (October 1953): 1–68.

PITZER, KENNETH S. "Role of the University in Basic Research," *Science,* 121 (June 3, 1955): 789–92.

REMMERS, HERMANN H. "The Expanding Role of Research," *North Central Association Quarterly,* 23 (April 1949): 369–76.

ROSANDER, A. C. "A Researcher Looks at Research," *Educational Record,* 21 (April 1940): 220–36.

RUARK, HENRY C. "Technology and Education," *Phi Delta Kappan,* 42 (June 1961): 387–92.

RUSSELL, DAVID H. and PARKER, J. CECIL. "Using Research to Point the Way in Curriculum Change," *Educational Leadership,* 12 (February 1955): 269–76.

RYANS, DAVID G. "Investigation of Teacher Characteristics," *Educational Record,* 34 (October 1953): 371–96.

———. "Measurement and Prediction of Teacher Effectiveness," in *1958 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1958.

———. *Characteristics of Teachers.* Washington, D.C.: American Council on Education, 1960.

SCATES, DOUGLAS E. "Needed: A New Pattern for Support for Educational Research," *Journal of Educational Research,* 39 (May 1946): 641–51.

———. "Fifty Years of Objective Measurement and Research in Education," *Journal of Educational Research,* 41 (December 1947): 241–64.

———. "Educational Research in the United States: Progress and Problems," *Educational Administration and Supervision,* 34 (November 1948): 385–411.

———. "On Our Research the Answers Hinge," *Phi Delta Kappan,* 30 (December 1948): 120–3.

———. "Recent Research and Current in Education," *Phi Delta Kappan* 30 (March 1949): 238–43.

———. "With the Researchers," See various issues of the *Journal of Teacher Education.*

SELLTIZ, CLAIRE and COOK, STUART W. "Can Research in Social Science Be Both Socially Useful and Scientifically Meaningful?" *American Sociological Review,* 13 (August 1948): 454–9.

SHANNON, JOHN R. "Teachers and Doers in Educational Research," *Journal of Educational Research,* 43 (October 1949): 142–5.

SHORES, J. HARLAN. *Critical Review of the Research on Elementary School Curriculum Organization,* 1898–1949. Urbana: Bureau of Research and Service, College of Education, University of Illinois, 1949.

SHUMSKY, ABRAHAM. "Cooperation in Action Research: A Rationale," *Journal of Educational Sociology,* 30 (December 1956): 180–5.

———. "Teachers Explore Their Attitudes Toward Research," *Educational Research Bulletin,* 37 (February 1958): 31–8.

SMITH, HERBERT A., *et al.* "Third Annual Review of Research in Science Teaching," *Science Education,* 39 (December 1955): 335–71.

SPENCE, RALPH B. "Research and Practice: The Case for Impure Research," in *Growing Points in Educational Research.* Washington, D.C.: American Educational Research Association, 1949, pp. 297–302.

STODDARD, GEORGE D. "Educational Research Lacks Impact; It Avoids Controversies and Human Values," *Nation's Schools,* 49 (May 1952): 44–6.

SYMONDS, PERCIVAL M. "Research Checklist in Educational Psychology," *Journal of Educational Psychology,* 47 (February 1956): 100–9.

———. "The Organization of Educational Research in the United

"States," *Harvard Educational Review*, 27 (Summer 1957): 159–67.

Tiedeman, David V. and Cogan, Morris L. "New Horizons in Educational Research," *Phi Delta Kappan*, 39 (March 1958): 286–91.

Traxler, Arthur E. "Research in Reading in the United States," *Journal of Educational Research*, 42 (March 1949): 481–99.

Traxler, Arthur E. and Jungeblut, Ann. "Research in Reading During Another Four Years: Summary and Bibliography," *Educational Record Bulletin*, No. 75. New York: Educational Records Bureau, 1960.

Traxler, Arthur E. and Townsend, Agatha. "Another Five Years of Research in Reading: Summary and Bibliography," *Educational Records Bulletin*, No. 46. New York: Educational Records Bureau, 1946.

Turner, Richard L. and Fattu, Nicholas A. "Skill in Teaching, A Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research," *Bulletin of the School of Education (Indiana University)*, 36, No. 3, 1960.

Tyler, Ralph W. "Trends in Teaching—How Research is Affecting Our Understanding of the Learning Process," *School Review*, 59 (May 1951): 263–72.

Tyler, Ralph W., *et al. Graduate Study in Education*. 50th Yearbook, Pt. 1. National Society for the Study of Education. Chicago: University of Chicago Press, 1951.

Walker, Fred L. "Blueprint for Knowledge," *Scientific Monthly*, 72 (February 1951): 90–101.

Wann, Kenneth D. "Teachers as Researchers," *Educational Leadership*, 9 (May 1952): 489–95.

Watters, William A. "Annotated Bibliography of Publications Related to Teacher Evaluation," *Journal of Experimental Education*, 22 (June 1954): 351–67.

Weaver, J. Fred (ed.). "A Bibliography of Selected Summaries and Critical Discussions of Research in Elementary-School Mathematics," *Arithmetic Teacher*, 7 (November 1960): 364–6.

Weitz, Henry. "Education is a Science," *Educational Forum*, 20 (March 1956): 293–301.

Wolfle, Dael, *et al.* "Standards for Appraising Psychological Research," *American Psychologist*, 4 (August 1949): 320–8.

Wrightstone, J. Wayne. "Research Action Programs for Research Bureaus," *Journal of Educational Research*, 42 (April 1949): 623–9.

———. "Influence of Research on Instruction in Arithmetic," *Mathematics Teacher*, 45 (March 1952): 187–92.

The promise of excellence in education rests on the willingness of the nation to support a comprehensive program of educational research and development to improve schools.

LINDLEY J. STILES

15 Significant Research Studies

The number of investigations conducted in the field of education is obviously large. An even larger number having a direct bearing on education have been conducted in related fields. The present chapter attempts to bring to the attention of the student a handful of classic studies of significance to educators; they should be of interest from the standpoint of both content and research design.

Space limitations have permitted the choice of only a few of the many studies of sufficient significance to warrant discussion here. The student is urged to check for additional titles in such sources as the *Encyclopedia of Educational Research*, *Review of Educational Research*, *Education Index* and the various professional journals, *Dissertation Abstracts*, the reports of the *Cooperative Research Project* of the U.S. Office of Education,[1] as well as more specialized sources such as Garrett's *Great Experiments in Psychology*,[2] the reports of the Kellogg Co-

[1] United States Office of Education, *Cooperative Research Projects* (Washington, D.C.: Government Printing Office, 1957–date).

[2] Henry E. Garrett, *Great Experiments* (3rd ed.; New York: Appleton-Century-Crofts, 1951).

operative Program in Educational Administration,[3,4] and textbooks in the various areas of educational specialization.

It has also been necessary to limit discussion of the studies listed to a bare orientation, which, in some cases at least, may not do them justice. More adequate treatment is to be found in the references cited; in all cases, the student is urged to consult the original sources for a more adequate grasp of the specific nature of the study.

THE MEASUREMENT OF INTELLIGENCE

ALFRED BINET

Of particular importance to educators is the well-known derivation of the first intelligence test by Alfred Binet who, in 1904, at the request of the French Ministry of Public Instruction, headed a commission to investigate the problem of retardation in the Paris schools. Realizing the relationship of intelligence to academic progress, he saw the need for a test to appraise intelligence or, more specifically, a test to locate those so mentally inadequate as to necessitate special care. In 1905, with the help of Theophile Simon, Binet published what might be considered the first test of general intelligence. This first attempt was revised in 1908 and again in 1911.

Departing from the then current emphasis on the atomistic measurement of narrow aspects of personality—rote memory, accuracy of perception, attention span, sensory discrimination, and so on—which, by this time, had been shown to be relatively sterile, Binet oriented himself toward measures of general intelligence with particular emphasis on the higher mental processes displayed in reasoning, imagination, judgment, attention, adaptability, and common sense. Putting together a number of items in rough order of difficulty, he attempted to allocate these to different age levels on the basis of the actual performance of children of different ages. In his 1908 revision, he arranged his items into age levels and coined the term "mental age." In his 1911 revision, which included age 3 to the

[3] Herold C. Hunt and Oliver R. Gibson, "CPEA, the Grand Design," *Nation's Schools,* 60 (October 1957) : 51–4.

[4] Hollis A. Moore, *Studies in School Administration: A Report of the C.P.E.A.* (Washington, D.C.: American Association of School Administrators, 1957).

adult level, he attempted to provide a variety of problem situations and to eliminate those items in which the factor of the specific experiences pecular to one child might feature too prominently.

The Binet tests were first introduced in America by Goddard who, in 1908, translated Binet's scale and extended and modified it for use in his work at the Vineland Training School. The most notable of the many revisions of the Binet scale in America is the Stanford revision by Terman (1916), and the later revisions by Terman and Merrill (1937, 1960). Terman introduced the concept of IQ in his 1916 revision and, of course, incorporated a number of other improvements and extensions. Nevertheless, the various editions of the Stanford-Binet are based on essentially the same theoretical conceptions and general arrangement as the original Binet test. The 1937 revision has for years been considered a standard in the area of intelligence testing; it is likely that the 1960 revision obtained by combining forms L and M of the 1937 scale will enjoy the same popularity.

(Another milestone in the area of the measurement of intelligence is the derivation of the Army Alpha and Army Beta tests for the classification of soldiers in World War I. The Army General Classification Test of World War II was, in a sense, a revision of the Army Alpha. In addition, a vast array of group intelligence tests have been devised.)

Binet's work is of particular interest because it broke away from the futile approach to the measurement of intelligence used up to that point and set the pattern which is still the basis of the bulk of current tests of intelligence the world over. Binet's contribution to the field of education is best appraised through an appreciation of the contributions which the measurement of the intelligence of thousands of youngsters the world over makes to the better calibration of curricular material and instruction to their level of understanding, their more adequate vocational orientation, and the other possibilities which cause modern educators to recognize intelligence tests as an indispensable tool. They serve a primary function at the college level in the screening of applicants. They are also used in industry and a number of other settings.

REFERENCES

BINET, ALFRED, and SIMON, THEOPHILE H. *The Development of Intelligence in Children* (trans. by E. S. Kite). Baltimore: Williams & Wilkins, 1916.

TERMAN, LEWIS M. and MERRILL, MAUD A. *Measuring Intelligence.* Boston: Houghton Mifflin, 1937.

———. *Stanford-Binet Intelligence Scale; Manual for the 3rd Revision,* Form L-M. New York: Houghton-Mifflin, 1960.

YERKES, ROBERT M. (ed.). *Psychological Examining in the United States Army.* Memoirs National Academy of Science, No. 15, 1921.

YOAKUM, CLARENCE S. and YERKES, ROBERT M. *Army Mental Tests.* New York: Holt, 1920.

THE HARVARD GROWTH STUDY

WALTER F. DEARBORN
and JOHN W. MCROTHNEY

This study was inaugurated in the fall of 1922 at the Psycho-Educational Clinic of the Harvard Graduate School of Education and continued for 12 years, during which approximately 3500 children entering first grade in the metropolitan area of Boston were examined annually from the first grade through adolescence. The study is a comprehensive longitudinal composite of a number of sub-studies aimed at getting a greater understanding of the general nature of growth or, more specifically, of differences in growth associated with individual variations, age, maturity, sex, and ethnic differences; of the nature and results of abnormal growth; of the relationship of physical growth to abnormalities in behavior; and of the relationship between mental and physical growth.

The results of the study led to the following conclusions (among others):

1. Physical and mental growth are essentially individual affairs; no two cases are the same; variability rather than consistency in growth is the rule; and comparison with average status has little value in the study of the development of individuals.

2. The relationship between physical measurement and mental measurement is so low that the knowledge of one does not enable us to predict the other.

3. Prediction of growth at various ages—except for group averages—is extremely hazardous, but it is particularly so during the period of adolescence.

4. The rate of development during the pre-pubescent growth spurt bears no significant relationship to learning of school material during this period. The rapid growth at adolescence need no longer be offered as an excuse for the slump in school performance during that period.

5. Classification of individuals into body types cannot be done with any substantial degree of accuracy.

6. An effective method of predicting body weight has been developed which takes into consideration the relative contributions of various bodily dimensions in determining weight.

7. The pre-pubescent growth spurt has been discovered to be much more abrupt than cross-sectional studies had led the authors to believe.

8. The timing of the pre-pubescent growth spurt is closely related to the advent of puberty.

9. Average differences between sex, age and ethnic groups are much less important than the individual variations found within each group.

The study has clarified many aspects of growth during the childhood and adolescent periods. It has probably made its greatest contribution in rejecting false notions about the relationship between physical and mental growth and in pointing out the highly individualistic nature of growth and the consequent limitations in the applicability of group norms to the individual. It should be of particular interest to teachers of the junior high school.

The authors conducted a companion study of the relationship of anthropological, physical, sociological, psychological, educational, and economic factors to employment and unemployment among young people. Their investigation of the employment status of 1360, out of a representative sample of 1541, subjects selected from the files of the twelve-year growth study just reviewed revealed essentially negative results. They found no relationship between the employment status of youth and chronological age, high-school attendance, absence and tardiness from school, school marks, IQ, attitude toward education, skeletal development, anthropological measurements, and

other aspects of growth. They did discover a relationship between employment and schooling beyond the high school. The authors point out that, while some of the findings may seem strange on the surface—for example, the apparent nonsignificance of academic grades and IQ in employment—it must be remembered that these aspects are frequently not appraised as part of the employee-selection procedures. It must also be remembered that the study was conducted in a period of economic depression.

REFERENCES

Cornell, Ethel L. and Armstrong, Charles M. "Forms of Mental Growth Patterns Revealed by Reanalysis of the Harvard Growth Data," *Child Development* 26 (September 1955): 169–204.

Dearborn, Walter F. and Rothney, John W. M. *Predicting the Child's Development.* Boston: Independent Press, 1941.

———. *Scholastic, Economic, and Social Backgrounds of Unemployed Youths.* Cambridge: Harvard University Press, 1938.

THE EIGHT YEAR STUDY OF THE PROGRESSIVE EDUCATION ASSOCIATION

The Eight Year Study was undertaken under the auspices of the Progressive Education Association for the purpose of obtaining dependable evidence regarding the relationship between the pattern of the high-school program and college success. It has long been the contention of some educators that if college-entrance requirements, which have, in large measure, governed the curriculum of the high school, were to be abandoned, secondary schools could improve their curricular offerings to the benefit of their students. The opposite viewpoint is that if these requirements were to be abandoned, chaos would result. The Eight Year Study put this notion to a test. In 1930, a committee of twenty-six, exploring the possibilities of better co-ordination between the high school and the college curriculum, felt that the high-school curriculum was too traditional to meet current student needs. They pointed particularly to the high school's neglect of its responsibility to those students for whom high school constitutes terminal education. They won-

dered further if such a curriculum was necessary or even helpful for college success.

The study lasted from 1933 to 1941 during the course of which thirty high schools—private and public—in various sections of the country that had indicated a willingness to liberalize their graduation requirements were allowed to make whatever changes they felt desirable. They received the assurance of some thirty colleges and universities that their graduates would be accepted on the recommendation of the principal without reference to the usual college-entrance requirements in Carnegie units. The schools were selected on the basis of their willingness to make exploratory curricular changes and the general competence of the staff to make such changes effective. The curricular changes made were varied in keeping with local needs and local facilities; there is no easy way to describe these except to say that they were indicative of the desire of the sponsoring school to replace inert subject matter by content more alive and pertinent from the standpoint of the problems of youth in modern civilization, and to give emphasis to the concept of education for the purpose of promoting the student's overall growth.

With respect to school-college co-ordination, a commission headed by Herbert E. Hawkes (of Columbia) investigated the progress in college of the graduates of the liberalized high schools. Since a large majority of the graduates of the high schools had enrolled in twenty-five colleges, the investigation concerned itself with the 1475 students enrolled therein. Each of these was carefully matched with a student graduating from a traditional curriculum on the basis of scholarship, age, sex, race, and home background. The results showed the graduates of the thirty schools to be on par with those of the traditional schools in the fundamentals but considerably superior in ability to reason critically, to apply what they knew, and to integrate their experiences. They also tended to be superior in co-operation, self-confidence, sociability, effectiveness of expression, interests, and creativity—that is, in the functionality of their learnings. The study also showed that the graduates from the schools that had departed most from the traditional curriculum did better than the graduates of those schools

which had made lesser changes. It would appear that the students were not handicapped with respect to college achievement by their unorthodox curriculum; in fact, departure from the traditional curriculum seemed to improve rather than to lessen their chances of success. It also pointed to the wisdom of relying on the judgment of the secondary school as to what constitutes adequate preparation for college. The study is particularly noteworthy for the ingenuity and the comprehensiveness of the instruments which were devised by the evaluation committee, under the direction of Ralph Tyler, for the purpose of evaluating student progress.

A similar series of studies was conducted by Wrightstone who compared traditional and "experimental" school practices in New York City. Using a matched-pair design over a six-year period, Wrightstone found the comparisons to favor the experimental group in all instances, particularly from the standpoint of social adequacy and critical thinking. However, despite the findings of these two studies, most colleges still subscribe to the traditional Carnegie unit entrance requirements. It might be a profitable exercise to relate the findings of this study to that of the Learned and Wood study of Pennsylvania high schools and colleges to be discussed later in this chapter.

REFERENCES

AIKIN, WILFORD M. *The Story of the Eight Year Study*. New York: Harper, 1942.

AMERICAN EDUCATION FELLOWSHIP (Progressive Education Assn.). *Thirty Schools Tell Their Story*. Adventures in American Education Series, Vol. 5. New York: Harper, 1943.

CHAMBERLIN, CHARLES D., *et al. Did They Succeed in College?* New York: Harper, 1942.

GILES, HARRY H., *et al. Exploring the Curriculum*. New York: Harper, 1942.

SMITH, EUGENE R. and TYLER, RALPH W. *Appraising and Recording Student Progress*. New York: Harper, 1942.

WRIGHTSTONE, J. WAYNE. *Appraisal of Newer Practices in Selected Public Schools*. New York: Teachers College, Columbia University, 1935.

———. *Appraisal of Experimental High School Practices*. New York: Teachers College, Columbia University, 1936.

———. *Appraisal of Newer Elementary School Practices*. New York: Teachers College, Columbia University, 1938.

PROJECT TALENT

JOHN C. FLANAGAN

Project Talent, currently under way, is an attempt to make a national inventory of the talents (aptitudes and abilities) of the students in the nation's secondary schools. The project is being conducted jointly by the University of Pittsburgh and the American Institute for Research, supported by funds from the Cooperative Research Program of the U.S. Office of Education with assistance from the National Institute of Mental Health, the Office of Naval Research, and the National Science Foundation.

In March, 1960, a stratified sample of 440,000 students in 1353 secondary schools in all parts of the country were given a two-day battery of tests covering such areas as common information, English, reading comprehension, memory for words and sentences, arithmetic computation, arithmetic reasoning, mathematics, abstract reasoning, mechanical reasoning, and creativity. Information on the student's background and plans, including his experiences, his study habits, his family background and so on, also was obtained. In addition, an attempt was made to measure the student's interest in various kinds of occupations and activities, and a further check was made to gather information about the kind of activities in which he actually engaged. Detailed information was collected concerning the guidance and counseling programs, the type of curriculum, and other educational practices, of the various schools of the nation, with a view to determining what may be expected from students currently attending high schools. Plans are to conduct follow-up studies of the original sample one, five, ten, and twenty years after graduation from high school and to relate this information to the data collected in 1960.

The ultimate goal of the study is to provide information that will lead to improved educational practices and policies in order to assist students in acquiring educational experiences which will lead them toward a realization of their full potential. It is realized that there are too many youngsters with special talents and a great deal of promise who never develop these talents; it is hoped that the project will identify these potentially talented people and enable them to make better use of

their particular patterns of aptitude. It is hoped that the study will contribute to improving the whole process of identifying, developing, and utilizing the talents of the nation's young people. The project probably represents the most comprehensive educational survey of all times; the feasibility of the study is, of course, a tribute to modern electronics, without which the conduct of a study of such magnitude obviously would have been out of the question.

(Of a somewhat similar nature is the Career Pattern Study conducted at the Horace Mann–Lincoln Institute of School Experimentation in an attempt to conceptualize the field of vocational development.)

REFERENCES

FLANAGAN, JOHN C. *Project Talent, 1. Design for the Study of American Youth.* New York: Houghton-Mifflin, 1962.

ORR, DAVID B. "Project Talent; A National Inventory of Aptitudes and Abilities," *Phi Delta Kappan,* 42 (March, 1961) : 237–43.

SUPER, DONALD E. and OVERSTREET, PHOEBE L. *The Vocational Maturity of Ninth-Grade Boys.* Career Pattern Study, Monog. No. 2. New York: Teachers College, Columbia University, 1960.

GENETIC STUDIES

ARNOLD GESELL

Gesell's work extends over half a century, from the time he entered the Yale Clinic of Child Development in 1911 to his death in 1961, during the course of which twenty-five publications have been made under his authorship and that of his associates. His investigations, sponsored by grants from the Rockefeller and Carnegie Foundations, cover such aspects as motor and physical growth and the development of emotional expression, philosophic outlook, adaptive behavior, language, interpersonal relationships, and personal-social behavior during the period from infancy to age 16.

These investigations are reported in three major publications: 1. *The First Five Years of Life* (1940) ; 2. *The Child from Five to Ten* (1946) ; and 3. *Youth: The Years from Ten to Sixteen* (1956) . Each of the investigations was essentially a longitudinal study of the same children;* the children in the

In some of the studies, the samples at different age levels were somewhat overlapping rather than completely longitudinal but, in all instances the study extended over a period of time, so that most of the cases were examined several times during the period of the study.

first study were observed at 4, 16, 28, 40, 52, and 80 weeks and at 2, 3, 4, and 5 years. The subjects of the second investigation were examined at 5, 5½, 6, 7, 8, 9, and 10. Many of these children had attended the nursery of the clinic and had been experimental subjects for some previous studies. Gesell's findings are reported in detailed description of growth patterns and norms of physical, mental, and personal-social development. His reports are particularly thorough from the standpoint of tables, charts, and illustrations.

In all cases, the emphasis has been on the maturational aspects, and Gesell has been criticized for minimizing the influence of environmental factors on development. Some of his critics have pointed out that his developmental norms suggest that growth is strictly maturational and "scheduled" so that at a certain age a child does a certain thing, at the next age level he does other things, regardless of the environmental conditions. Gesell has maintained this position in the face of the trend towards environmentalism sponsored by Watson and others. His study with Thompson comparing the development of identical twins under different conditions of practice attempted to point out that environment and practice are relatively ineffective in promoting growth unless and until the required maturational readiness is present. He found that the untrained twin caught up in his development after a period of maturation.

Gesell's studies have also been criticized for failure to use unselected populations. His subjects were in the main from the upper socio-economic and cultural strata representing the more stable families in New Haven (Conn.) and its suburbs. (The average IQ of his sample for his report on youth, for example, was approximately 117.) This might tend to bias the results in view of the relatively convincing evidence of a differential developmental pattern for people of different socio-economic levels.

Nevertheless, Gesell's work constitutes a major contribution to the understanding of the child and his development. From the standpoint of research, his studies also represent the type of systematic, programmed research—involving an extensive staff of trained observers using a multiple approach of standardized test data, sequential examinations, clinical interviews, and observations supplemented by movie cameras, one

way screens, and other modern means—necessary for productive research and for the progress of education as a science. His work is probably among the best known to lay, as well as to professional, people interested in understanding and promoting maximum child growth; they are of particular interest as a guide to parents and teachers of the preschool and elementary grades.

REFERENCES

GESELL, ARNOLD. *Infancy and Human Growth.* New York: McMillan, 1928.

———. *The First Five Years of Life.* New York: Harper, 1940.

———. *Infant Behavior.* New York: McGraw-Hill, 1934.

GESELL, ARNOLD and GESELL, BEATRICE C. *The Normal Child and Primary Education.* Boston: Ginn, 1912.

GESELL, ARNOLD and ILG, FRANCES L. *The Child from Five to Ten.* New York: Harper, 1946.

GESELL, ARNOLD and THOMPSON, ELLEN. *Learning and Growth in Identical Twin Infants.* Genetic Psychology Monograph, 6, No. 1 (1929): 1–124.

GESELL, ARNOLD, *et al. Biographies of Child Development.* New York: Hoeber, 1939.

GESELL, ARNOLD, *et al. Youth: The Years from Ten to Sixteen.* New York: Harper, 1956.

ILG, FRANCES and AMES, LOUISE B. *The Gesell Institute of Child Behavior.* New York: Bell Publishing Co., 1955.

STOLZ, LOIS M. "The Gesell Institute and its Latest Study," *Contemporary Psychology,* 3 (January, 1958): 10–14.

THE CHARACTER EDUCATION INQUIRY

HUGH HARTSHORNE
and MARK A. MAY

The Character Education Inquiry was undertaken in 1924 under the auspices of the Religious Education Association in an attempt to evaluate the result of moral education. The investigation had two major purposes: the study of deception, and the development of instruments to appraise moral knowledge and attitude. The study is reported in three volumes: 1. *Studies in Deceit* (1928); 2. *Studies in Service and Self-Control* (1929); and 3. *Studies in the Organization of Character* (1930), the best known of which is the first. Originally planned for a three-year period, it was extended to five years, a good

portion of which was spent in the development of reliable and valid procedures for studying character.

In the study of deceit, over ten thousand students from both public and private schools from grades one through twelve and from all varieties of socio-economic, cultural, ethnic, intellectual, occupational, community, and religious backgrounds were placed in semi-laboratory situations where they could cheat, lie, steal, or refrain from so doing. These situations were kept as natural as possible in order to appraise the self and its integration with group standards and expectations—that is, in order to see the individual under conditions of normal social interaction. Twenty-nine test situations, many of considerable ingenuity, were devised to measure the extent of deceit. Twenty-two of these "deception tests" involved ordinary classroom situations; four took place in an athletic setting, two at parties, and one involved work done at home. There were also two "lying" and two "stealing" tests. It was hoped that these tests would provide a relatively complete picture of the individual's tendencies to deceive.

The results revealed that children engage in a considerable amount of deceit, and that deceit is associated with such personal traits as dullness, retardation, emotional instability, low academic achievement, socio-economic and cultural limitations, certain national, racial, and religious groupings, disciplinary problems in school, and attendance at movies. The investigation, for example, suggested that deception runs in families in much the same way as intelligence or eye-color. This, of course, simply implies concomitance; in no way is there an implication of causation or inheritance. Deception seems to be affected by social interaction; the behavior of his close friends and associates, for instance, were more basic indicators of whether a child would cheat than his associations with adults. Deception in a classroom was at a minimum when an atmosphere of goodwill and co-operation existed between the teacher and pupils and a general high level of morale existed in the school. Attendance at Sunday School or membership in scouting and other clubs oriented toward the teaching of integrity and honesty did not seem to have much influence, if any; in fact, these children actually appeared less honest than average.

It was found that children were not consistent in their be-

havior; a given child might be honest under one set of circumstances but not under another. The findings seemed to point to honesty as a conglomeration of specific acts governed largely by the specific situation in which the child found himself. Generally, the most common extraneous motive leading to cheating was a desire to do well in class. On this basis, it would seem that social control of deceit is best approached through manipulating the situation in such a way as to make deceit unnecessary. This generalization runs counter to the integrative and directive nature of the self-concept as presented by Snygg and Combs, and appears illogical from a psychological point of view in that it makes the individual's behavior essentially haphazard, chaotic, and otherwise mechanical, rather than organismic. An argument against such a conception of character has been presented elsewhere.

The authors acknowledge that they had measured deceit (or conduct) rather than character—that is, that the tests used were measures of deception, helpfulness, co-operation, inhibition, and persistence, all of which are aspects of behavior which comprise character only when they are integrated into a functional mass. The inquiry constitutes a pioneer study in an important area; the findings have obvious implications for character education as sponsored both by the public schools and by specialized agencies such as the church, youth agencies, and parochial schools. To the extent that the school accepts character formation as a major goal, it must be concerned with the effectiveness of its efforts in this direction. Obviously the findings also have broad sociological implications.

REFERENCES

HARTSHORNE, HUGH. "Sociological Implications of the Character Education Inquiry," *American Journal of Sociology,* 36: 251–62, September, 1930.

———. "Recent Improvements in Devices for Rating Character," *Journal of Social Psychology,* 1 (January, 1930): 66–77.

HARTSHORNE, HUGH and MAY, MARK A. 1. *Studies in Deceit* (1928); 2. *Studies in Service and Self-Control* (1929); and 3. *Studies in the Organization of Character* (1930). New York: MacMillan.

MOULY, GEORGE J. *Psychology for Effective Teaching.* New York: Holt, 1960.

SNYGG, DONALD, and COMBS, ARTHUR W. *Individual Behavior.* New York: Harper, 1949, 1959.

THE HAWTHORNE STUDIES

THE HARVARD SCHOOL OF BUSINESS

This investigation was a series of studies conducted at the Hawthorne (Chicago) plant of Western Electric Co., a subsidiary of Bell Telephone Co., the manufacturer of telephone equipment for American Telephone and Telegraph Co. Originally designed as a one-year study of the effects of fatigue and monotony on worker output, the investigation extends from the mid-twenties to its present integration into the standard operation of the company. The major report, published in 1939, covers the first twelve years of the study.

The results of the various sub-studies reveal a pattern of increase in worker productivity attending almost any and all changes in working conditions. In the study of illumination, for example, in which the production of four groups of supervisors, coil winders, and relay assemblers was compared under four different lighting conditions, output increased with an increase in lighting, but it also increased when the illumination was brought back to its original level and even when it was reduced to a mere three foot-candles. In fact, two girls who volunteered to work under lighting equivalent to that of ordinary moonlight were as efficient, reported no eye-strain, and showed less fatigue than when working under more normal lighting. Output also increased when light bulbs were simply exchanged for other bulbs of the same wattage, after the workers had been led to believe that a change in light intensity was being made.

In order to exert greater control over the situation, the investigators separated five girls assembling relays from the regular working force and placed them in a special test room where changes could be made without disturbing the operation of the rest of the plant. A major part of the investigation concerned the output of this group under a series of experimental conditions relating to lighting, rest pauses, length of work-day and work-week, wages and pay incentives, and other aspects of their working conditions. Production increased in nearly all cases, the increase reaching a maximum of about 30 percent above pre-experimental standards. It rose, for example, when the work-week was shortened and when rest pauses

were introduced, and it rose again when both factors were returned to their original levels.

The results led the investigators to the conclusion that social factors connected with the experimental changes—rather than the changes themselves—were responsible for the increase in output. A major feature of the study was the status which the girls enjoyed. Their advice and reactions concerning the changes to be made in their working conditions were sought by the plant superintendent. They were made to feel that the management was interested in their welfare, and they were given both special attention and special privileges not available to workers in the regular department. The findings of the study suggested that the increased output resulted not so much from the improved lighting or the rest pauses, and so on as from the feeling of status and morale. In the pay incentive study, for example, the output of the girls in the test room increased when bonuses were based on the production of the five-member team rather than on that of the total working force. This is understandable. But it also increased when their pay incentives were returned to a total-working-force basis.

The most significant aspect of the various changes that took place in the test room was the social transformation of the girls. There was not only a considerable increase in the morale and cohesiveness of the group but both grievances and absenteeism decreased sharply. As a result of the findings of the study, the study shifted from a consideration of the influence of the physical aspects of the working situation on production to an investigation of the more subtle and intangible factors of the psychology of personal and social adjustment in an industrial setting. A mass interviewing program in which over 20,000 interviews were held with employees of the plant bore out the investigators' position that the employee's social status in his work group is a major cause of employee concern and complaint. The size and location of his desk or work position, for example, constitutes a status symbol frequently of greater concern to him than his salary.

Another phase of the study revealed that it is the standards of production of the group, rather than his personal goals, that determine employee performance. Even with incentive

pay, individual and group output is, to a large extent, dictated by the worker's fear of being a "rate buster" if he turns out too much, a "chiseler" if he turns out too little, and, of course, a "squealer" if he reports his fellow-workers. These considerations set definite limits to the individual's production performance and even to his personal relations with "management." Furthermore, his need for group acceptance and status forces him to give group standards precedence over the more tangible and concrete company-employee relationships.

The studies are liable to certain criticism: for example, the investigation revolved around a very small group of volunteers and/or specially selected individuals. In fact, two of the five girls in the original test room situation were actually replaced because their low production created friction with the other three. Nevertheless, the investigation has made a significant contribution to our understanding of the complexity of the human relations problem in industrial production. More specifically, it has pointed out that wages, hours of work, working conditions, and the other aspects of the working situation are primarily carriers of social values concerning the individual's position or status among his immediate fellow-workers and in the company as a whole. It has shown that the worker's attitudes are basic determinants of production—affecting both individual and group effort—and, though many of his attitudes are perhaps irrational when viewed objectively, an understanding—and where necessary, a reorientation—of these attitudes is essential to effective employee management.

The study represents an honest and concerted effort to understand workers as individuals, and it makes industrial efficiency not only a mechanistic problem in engineering but also a problem of personal and group dynamics. The study has undoubtedly had considerable influence in the evolution of the pattern of modern industrial management-employee relationships. Not only has it helped to introduce a new concept of "industrial psychology" with emphasis on leadership, democratic supervision, and human relations, but it has also set the pattern for similar studies of this important area. The investigation, although conducted in an industrial rather than an academic setting, has direct bearing on classroom performance at both the

individual and the group level. The modern emphasis in education on motivation, attitudes, and group dynamics is consistent with the findings of this study.

REFERENCES

MAYO, E. *The Human Problems of an Industrial Civilization.* New York: Macmillan, 1933.

ROETHLISBERGER, F. J. *Management and Morale.* Cambridge: Harvard University Press, 1941.

ROETHLISBERGER, F. J. and DICKSON, WILLIAM J. *Management and the Worker.* Cambridge: Harvard University Press, 1939.

THE STUDENT AND HIS KNOWLEDGE

WILLIAM S. LEARNED
and BEN D. WOOD

This study was conducted in Pennsylvania as part of a sequence of inquiries financed by the Carnegie Corporation. The study, which involved the administration of a comprehensive testing program to 45,000 high-school and college students, placed primary emphasis upon knowledge with an underlying hope of ending the rule of the college credit as the measure of academic adequacy and progress. While realizing the importance of such supplementary traits as character, attitude, and social efficiency, the authors were of the opinion that the basic criterion of college acceptability and college progress is still knowledge or, more specifically, permanent and available knowledge, which is sufficiently defined and digested that it is readily available when needed so that it can serve as the basis for producing more advanced knowledge.

Three successive examinations were given: 1. the testing of high-school seniors in 1928; 2. the testing of the same group at the end of the sophomore year in college (1930); and 3. a third testing of the same group in their senior year in college (1932). The tests used were devised especially for the inquiry and involved such phases as English, mathematics, history and social studies, natural sciences, and so on. The examination, a copy of which is available in the appendix of the report, required twelve hours of testing at the senior-high-school level and eight hours for the college groups.

The most significant aspect of the findings was the great variability in performance among the participants; there was a

great variation among the institutions and a "much more striking" variation among the students of any one institution. In one phase of the study, tests were given to 5747 college sophomores attending forty-nine different institutions, to 3720 seniors, and to 1503 high-school seniors. From the many significant comparisons presented, the authors point out that school status (defined by the time spent and the courses passed in high school or college) has little relationship to any definite body of ideas, understood and available as a result of "education." The results showed that 28 percent of the college seniors did less well than the average sophomore and that nearly 10 percent did less well than the average high-school senior; conversely, 22 percent of the high-school seniors surpassed the average college sophomore and 10 percent of the high-school seniors surpassed the average college senior. Stated differently, the scores among sophomores ranged all the way from what might be considered inferior high-school achievement to a degree of excellence which is attained only by the best ten percent of college seniors—which, in fact, the authors point out, is perhaps above the average of faculty groups "if our experience on the earlier examination (1928) may be trusted."

The authors further point out that if, instead of graduating seniors simply on the basis that they had been on campus four years and have accumulated the required number of credits, graduation had been based on knowledge as revealed by test performance, the graduating class of 1932 would have consisted of the top 28 percent of the seniors, the top 21 percent of the juniors, the top 19 percent of the sophomores, and the top 15 percent of the freshmen. This hypothetical graduation class would have far surpassed in knowledge the class of seniors that did graduate and, of course, would have been nearly three years younger. It was also noted that only two-thirds of the freshmen who made the grade in this hypothetical graduating class were still in attendance at the sophomore level, and that two thirds of those actually tested lower than they had tested as freshmen. This, the authors interpret as evidence that, as presently organized, courses lead to the accumulation of college credits rather than the accumulation of knowledge. In the high-school study (1928) covering over 26,000 high school seniors, it was found that about 25 percent of the non-college prep stu-

dents scored above the average of the college-bound group and vice versa.

A phase of the investigation concerned the comparison of the test performance of college seniors planning to teach and that of high-school seniors. As might be expected, there was a substantial overlap between the two distributions, not only in knowledge of subject matter but even in basic vocabulary. An even more pointed comparison showed that 50 percent of the high-school students specializing in science had higher science scores than nearly 40 percent of the college seniors planning to teach science; conversely, 17 percent of the college teacher specialists had lower scores than 31 percent of the corresponding high-school students. In other words, many high-school students actually surpassed their prospective teachers with respect to knowledge in their own fields.

The authors interpret the results as an indication of the unsuitability of the present college curricular organization, where the goal is the passing of examinations and the accumulation of credits, each tied to specific courses, rather than to the accumulation of knowledge from broad and varied sources. (They present a number of recommendations which need to be read in the original to be appreciated.)

The study is, of course, dated; yet a repetition would, undoubtedly, reveal essentially the same conditions prevalent today. Whether this represents the lamentable condition Learned and Wood imply, and whether college credits and college degrees should be awarded on the basis of academic competence, however acquired, is a question of philosophy and beyond the scope of research *per se*.

REFERENCE

LEARNED, WILLIAM S. and WOOD, BEN D. *The Student and His Knowledge*. New York: Carnegie Foundation, 1938.

PATTERNS OF AGGRESSIVE BEHAVIOR IN EXPERIMENTALLY CREATED SOCIAL CLIMATES

KURT LEWIN

This is a report of a number of experiments conducted in the Child Welfare Research Station of the State University of Iowa dealing with the concept of group dynamics. In one of the

experiments, Lippitt organized two clubs of ten-year-old boys engaged in the activity of making masks: one of the clubs was governed in an autocratic, authoritarian manner while the other operated on a democratic basis. In a second experiment by White and Lippitt, four new clubs of ten-year-old boys, engaged in mask-making as well as mural-painting, soap-carving, model airplane construction, and so on, were organized on a voluntary basis, each under a different type of adult leadership; one group was governed democratically, the second had an autocratic leader, the third operated on a *laissez-faire* basis without adult leadership. Every six weeks, the groups changed leadership so that each of the groups had three different leaders in the five months of the experiment. The groups were equated with respect to teacher ratings on such items as socio-economic background, social behavior, leadership potential, interpersonal relations, intellectual status, physical status, and other personality characteristics. There were eleven meetings of each group: the democratic group met first and engaged in activities of their own (group) choosing. In order to maintain equivalence of the tasks, the autocratic group, in its meeting two days later, was assigned the activities which the democratic group had selected. The *laissez-faire* group was simply left on its own.

In the autocratic group, the policy was determined by the leader, and one step—what should be done, by whom, and with whom—was dictated at a time so that the future steps were always uncertain to a large degree. In general, the leader was aloof from the group—that is, impersonal rather than unfriendly. In the democratic group, all policies were determined by group discussion encouraged and assisted by the adult leader. The *laissez-faire* group was given complete freedom for group and individual decision; the materials were supplied, but the leader made it clear that he would provide information only when asked. He did not participate actively in any of the activities. It should be noted, however, that even in the autocratic group, participation was voluntary and the relationships were essentially congenial.

As the meetings progressed, the authoritarian club members developed a pattern of aggressive domination toward one another while their relationship to the leader became one of greater submission or of persistent demands for attention. The

authoritarian group was significantly more aggressive and hostile than the democratic group. There was scapegoating and two of the members ceased coming to the meetings. Interviews with the boys also showed complete agreement on the relative dislike of the autocratic leader, regardless of who he was. The aggressive pattern was even stronger in the *laissez-faire* group; this is probably best explained by the fact of the freer atmosphere which permitted aggressiveness to be shown. Aggressiveness was frequently controlled and suppressed in the autocratic group when the leader was present; it showed itself, however, when supervision was removed. It is also very likely that under conditions of autocratic control, apathy sets in, and the autocratic group was found to be rather dull, lifeless, submissive, repressed, and apathetic; there was little joking, smiling, freedom of movement, or freedom for initiating projects. In the second experiment, four of the five autocratic groups became rather apathetic, but apparently they were still hostile, as exhibited by aggressiveness toward one another when the leader left the room, by their expressed dislike for the autocratic leader, and by the general absence of smiling, joking, and so on.

The democratic atmosphere, on the other hand, produced more constructive suggestions, more frequent matter-of-fact behavior of member to member, greater individuality, and greater co-operation. The democratic group was more spontaneous and friendly; it was characterized by a great deal of "we-feeling" as opposed to the "I-feeling" to be found in the autocratic group.

It is also interesting to note that as two children were switched from one group to the other, each took on the characteristics of the group to which he was transferred. Similarly, as the groups were changed from autocratic to democratic leadership, the members assumed the pattern typical of the group to which they were assigned. It did, however, take somewhat longer for the autocratic group to adjust to democratic procedures than for the democratic group to adjust to autocratic control, suggesting that, while autocracy is imposed upon the individual, democracy has to be learned. It appears a fallacy to assume that, if left alone, individuals will form themselves naturally into democratic groups; chaos or a primitive pattern of organization through autocratic dominance by a few members is a more likely outcome.

Although this study was not conducted in a classroom set-

ting and does not involve ordinary academic learnings, it has very definite educational implications. Lewin concluded that the social climate in which a child lives is as important as the air he breathes, that the group to which the child belongs is the ground on which he stands, and that it is all-important to his security. He points out further that the success a teacher is likely to have in a classroom depends not only on his skill but also on the classroom atmosphere which he creates. This may be even more true for the intangible aspects of education. Actually, the democracy and autocracy which Lewin discusses are ideal autocracies and democracies, rather than what one might find in the field. It might be well to explore the possibilities of conducting a similar experiment under more normal conditions and in a more standard academic setting.

REFERENCES

LEWIN, KURT. "Experiments in Social Space," *Harvard Educational Review,* 9 (January, 1939): 21–32.

———. "The Dynamics of Group Action," *Educational Leadership,* 1 (January, 1943): 195–200.

LEWIN, KURT, *et al.* "Pattern of Aggressive Behavior in Experimentally Created Social Climates," *Journal of Social Psychology,* 10 (May, 1939): 271–99.

LIPPITT, RONALD. "An Experimental Study of the Effects of Democratic and Autocratic Group Atmospheres," *Studies in Topological and Vector Psychology,* 1. Iowa City: University of Iowa Press, 1940.

THE QUALITY OF GROUP DECISIONS AS INFLUENCED BY THE DISCUSSION LEADER

NORMAN R. F. MAIER

Research has shown that a decision which is arrived at collectively is more acceptable to a group than one which is imposed by someone in authority. It is also possible that group thinking may be superior to that of the individual, since the thinking of a number of individuals is combined in a group decision. On the other hand, it must be recognized that the supervisor or leader is very frequently more informed, with a richer background of experience, and that he can, therefore, make valuable contributions to the group's thinking. The question arises as to whether or not he should refrain from participating in order to have the group arrive at its own decisions

so that it will accept the decisions that are reached. Or, restated, it is a matter of whether or not a leader can make his contributions to the group without incurring group resistance to the implementation of his ideas.

This study consisted of presenting a group of college students with a problem situation picturing an industrial assembly line in which production was being delayed by one of the men who was not as competent as his fellow-workers. In one of the experiments, the leader was well-versed in democratic processes and personality dynamics. He did not furnish the solution but restricted his contributions to summarizing, encouraging, analyzing, interpreting, supplying information, and preventing hurt feelings. He did make suggestions concerning the solution in the way he asked the questions and his wording of the suggestions of others, but he did not make the solution obvious. In the second experiment, each of the subjects in the study took the place of one of the workers on the assembly line. The third experiment involved the use of untrained leaders who were merely given preliminary training in the nature of the problem, the procedures to be followed, and what might constitute an adequate solution. A control group discussed the problem without a discussion leader.

The results indicated that a group leader can greatly improve the quality of the group's thinking and that, generally, the competence of the leader determines the quality of the decision that is reached. There still remains the question as to whether or not the solution will be accepted since an inferior solution that is accepted and implemented may be more functional than a more adequate solution that is not accepted. The results indicated that the solution was accepted in the majority of the cases with both the untrained and the trained leaders. The experiment showed that a leader who is skilled and who has ideas can conduct a discussion to obtain a better decision than that obtained by a group working with a less skilled leader or alone. He can also attain a higher degree of acceptance than a less skilled leader. In fact, the quality of the decision which can be reached under skilled leadership very frequently increases the acceptability of the solution. However, even an unskilled leader, relying on basic democratic conference procedures, can apparently promote decisions both of quality and of

a high level of group acceptance—suggesting that a leader who has ability in solving technical problems need not sacrifice his ability in order to maintain group goodwill. It must, of course, be recognized that a great deal depends on the rapport the leader is able to establish with the group. It must also be noted that in this study even the untrained leader had the correct solution to the problem; this ideal situation may not prevail in a real-life setting.

The study should be of interest to educators who, in a major sense, operate as classroom leaders. It is especially important with the problem solving approach currently emphasized as a classroom procedure. Further, inasmuch as the teacher is likely to be an informed leader, his contributions to the classroom discussion, provided he operates in an atmosphere of co-operation, can probably lead to superior solutions without incurring the dangers of resistance in its implementation.

REFERENCE

MAIER, NORMAN R. F. "The Quality of Group Decisions as Influenced by the Discussion Leader," *Human Relations,* 3, No. 2 (1950): 155–74.

COOPERATION AND COMPETITION: AN EXPERIMENTAL STUDY IN MOTIVATION

J. BERNARD MALLER

A study of considerable importance to educators in view of the crucial nature of motivation in classroom performance is Maller's comparison of the relative motivational force of co-operation and competition. The study attempted to compare the performance in simple addition of children under conditions of co-operation or self-motivation (that is, appeal to desire for personal gain) with their performance under conditions of group motivation or appeal to desire for group gain. A secondary purpose was to discover the concomitant factors associated with either tendency. The major part of the study consisted of three basic experiments involving 814 children in grades five through eight in four schools of different socio-economic status. The overall study involved 1530 children from ten different schools.

The first experiment compared pupil performance in twelve sessions of work. In six of these, each child worked under

conditions of competition; in the remaining six he worked under conditions of co-operation. A control group worked without any particular incentive. The results favored competition over co-operation, and both competition and co-operation over the control situation. These differences were statistically significant; furthermore, the superiority of performance under conditions of competition over that under conditions of co-operation increased with the practice sessions. The greater motivational force of competition tended to be more pronounced among girls than boys. On the other hand, about one third of the students performed better under conditions of co-operation.

In the second experiment, each child was allowed to choose whether he would work for himself or for the group. The subjects chose to work for themselves in three quarters of the trials, and their performance again favored competition. Again, certain children reversed the trend in performance.

In order to identify the traits associated with competitiveness and co-operation, Maller selected the 200 most co-operative and the 200 most competitive members of his overall group. These he compared on the basis of reputation, behavior, intelligence, scholastic status, physical traits, social and moral traits, home environment, and so on. The attempt was not particularly successful; in general, competitiveness and co-operation appeared to be relatively independent of sex, age, scholastic standing, health, and nationality. Co-operation was found to be slightly correlated with intelligence, moral knowledge, and resistance to suggestion.

In the third experiment, Maller compared competition as an incentive with various forms of co-operation: teamwork (teams chosen by captains), partnership (co-operation in pairs), boys *versus* girls, arbitrary grouping (class divided arbitrarily into two groups by the experimenter), and overall grouping of the class as a whole. Once again, pupil performance favored competition on an overall basis, but performance under conditions of competition was found to be inferior to that motivated through boys-*versus*-girls co-operation.

It is to be noted that Maller's design in the second experiment places co-operation in a setting of competition—that is, he defines co-operation as members co-operating as a group in order to compete more advantageously against another group.

The superiority of co-operation over competition under conditions of sex rivalry (boys *versus* girls) must be considered in that light. He did not compare competition against co-operation with the latter devoid of obvious competition. Nor did he consider the factor of group cohesiveness. Also important is the fact that the task which Maller imposed was not a particularly co-operative activity in which individual success is dependent on group co-operation—that is, the task tended to lend itself readily to independent work—and, of course, the study pertains to American children whose up-bringing is possibly more competitive than that of certain other national groups. His results may also be a function of the age level. Maller suggests that the lack of practice in group co-operation—that is, the lack of training in working with a group for a common goal—precludes the formation of habits of co-operation and group loyalty, a possibility which may have been somewhat more true in the twenties than it is of the present organization of the school. Despite its possible limitations, Maller's study has interesting implications for our present society and for the school whose orientation at present is still essentially competitive.

REFERENCES

DEUTSCH, MORTON. "A Theory of Cooperation and Competition," *Human Relations*, 2 (April, 1949): 129–52.

MALLER, J. BERNARD. *Cooperation and Competition: An Experimental Study in Motivation.* Teachers College Contributions to Education, No. 384. New York: Columbia University, 1929.

MAY, MARK A., *et al. Memorandum on Research in Competition and Cooperation.* Bulletin, No. 25. New York: Social Science Research Council, 1937.

WHEN SHOULD CHILDREN BEGIN TO READ?

MABEL V. MORPHETT
and CARLETON WASHBURNE

The subjects of this important study, conducted in 1928–29, comprised the total first-grade enrollment of the Winnetka (Illinois) public school. ($n = 141$.) Beginning first-graders were tested by the Detroit First Grade Intelligence Test and the Stanford Revision of the Binet test. Reading proficiency was measured by sight-word lists and the Gray Standardized Oral Reading Check. All eight of the Winnetka first-grade teachers

co-operated in the study, but they were not told the mental age of their pupils.

The study attempted to relate the reading progress of first-graders to the level of their mental development. The results can be summarized as follows: when the Detroit test was used as the basis for determining intellectual status, the children who had a mental age of six years and six months made far greater progress than did those who were less mentally mature, and practically as satisfactory progress as children of a higher mental age. When the mental age was measured by the Stanford-Binet, the children with a mental age of six years and six months again made better progress in reading than did those of lesser mental maturity; however, they made somewhat less satisfactory progress than did those of greater mental ages. A repetition of the experiment in 1929–30 with different teachers and different children confirmed the earlier experiments in all its basic conclusions.

The study is frequently cited in support of the position that it takes a mental age of six years and six months in order to learn to read the way reading is taught in our schools today. It is felt that this mental age lessens the likelihood of failure and discouragement attending the child's attempt to read when he is not "ready," and that, correspondingly, it increases the effectiveness of the school.

The study should be related to that of Gates, conducted some eight years later, who attempted to determine the optimum mental age at which reading should be introduced. In contrast to the findings of Morphett and Washburne, Gates pointed out that the crucial mental age for reading varies with the material and the type of teaching; the mental age that is required for learning to read under one program or with one method may be entirely different from that required under other circumstances. Gates concluded that the determination of the optimum age for reading readiness is not as simple as Morphett and Washburne had suggested, and that it is possible for a child with a mental age of five to learn to read.

REFERENCES

GATES, ARTHUR I. "The Necessary Mental Age for Beginning Reading," *Elementary School Journal,* 37 (March, 1937): 497–508.

MORPHETT, MABEL V. and WASHBURNE, CARLETON. "When Should Children Begin to Read?" *Elementary School Journal,* 31 (March, 1931): 496–503.
WASHBURNE, CARLETON. "Ripeness," *Progressive Education,* 13 (February, 1936): 125–30.

SPELLING INQUIRY

JOSEPH M. RICE

Probably none of the pioneers in early educational research deserves more credit than Rice, whose investigation of spelling marks education's first attempt at the objective study of educational problems. Although experimentation in the physical sciences and lab experimentation anteceded his relatively crude investigation, his study of spelling, conducted in the early 1890's, constitutes the first attempt at educational field experimentation.

Born in 1857, Rice received his M.D. degree in 1881. After a brief practice of medicine, he went to Europe where he studied pedagogics and psychology at the major educational centers of Germany. After his return to America, he devoted his whole energy to the improvement of American educational practice which he felt dragged behind that of Europe and, particularly, of Germany.

As editor of *Forum,* he campaigned vigorously to improve school practice. As one of his first activities, he undertook a relatively comprehensive tour of American schools, visiting some twelve hundred teachers in the Eastern and Mid-Western states. He was particularly disappointed at the mechanical way in which learning took place, the emphasis on isolated facts, and the failure to relate education to pupil interest. The articles which he wrote in support of more effective teaching were largely ignored.

However, Rice was not so easily discouraged; he proceeded to collect evidence to prove his point. Among the problems current at that time was a movement toward extending the curriculum to include such subjects as home economics and manual training. This extension, Rice noted, was opposed by many—including many educators—who felt that any addition to the curriculum had to be made at the expense of the basic curriculum then in vogue. Rice, basing his views on what he felt was an inadequate use of school time, rejected the assumption that

the results to be obtained from the learning of any one subject was proportional to the time devoted to it. Choosing spelling as an area in which he felt much of the instruction was essentially lifeless and unprofitable, he devised a test which he administered to some 100,000 children. His results showed little relationship between the spelling gains noted and the class time spent on the subject; schools devoting ten to fifteen minutes a week to the subject achieved gains equal to those spending as much as an hour.

To say that Rice's findings were not particularly well received is an understatement. They were ignored wherever possible or disputed and discounted; they had relatively little effect on educational practice for at least a quarter century. It is only in retrospect that his study attains significance, for it marks the beginning of education's reliance on evidence in the solution of educational problems and, accordingly, represents a major step toward the modern view that educational problems must be settled by investigation, rather than by argumentation.

Rice's work was not limited to the investigation of spelling; as editor of *Forum* he published some twenty articles oriented toward the improvement of educational practice. His contributions are most adequately reported in his book, *Scientific Management in Education.* The rejection of his findings points to a common problem faced by research workers. Generally, the efforts of an outsider who uncovers weaknesses in the educational program are resented.

REFERENCES

RICE, JOSEPH M. "The Futility of the Spelling Grind," *Forum,* 23 (April and June, 1897): 163–72, 409–19.

———. *Scientific Management in Education.* New York: Hinds, Noble, and Eldredge, 1914.

SCATES, DOUGLAS E. "Fifty Years of Objective Measurement and Research in Education," *Journal of Educational Research,* 41 (December, 1947): 241–64.

LEVELS OF ASPIRATION IN ACADEMICALLY SUCCESSFUL AND UNSUCCESSFUL CHILDREN

PAULINE S. SEARS

The meaning which a task has for the individual must be considered from the standpoint of its relationship to the self-concept which he has built up over the years. Generally, the

individual attempts to compete with himself; he may have a need to perform well in order to maintain status, and he generally strives to make a good showing. In addition, he generally attempts to derive from the situation in which he finds himself some degree of social approval from persons whose appraisals he values. Sears' study was concerned with the effect which past success or failure has on the level of aspiration of ego-involved subjects who have experienced characteristically continued success or continued failure with a particular task, and who, therefore, have certain expectations relative to their ability to perform adequately. It was her hypothesis that his level of aspiration in the performance of a given task is a function of the success-failure status of the past experiences which a student associates with the task.

The subjects were children of the fourth through the sixth grades, and the tasks were those dealing with reading and arithmetic with which the children had already has some contact. The success group was composed of children who during their entire school life had shown evidence of success in academic subjects including reading and arithmetic. The failure group, on the other hand, had the opposite experience. A third group was made up of children who had had success experiences with reading and failure experiences with arithmetic. All the children selected were ego-involved in the sense of being interested in the quality of their performance. The three groups were equated on such variables as chronological age, mental age, and sex.

After a preliminary trial, or neutral, session, the success group was advised that it had done exceedingly well and each member was told: "You did the first test in so many seconds: what are you going to try to do it in this time?" The failure group, on the other hand, was rebuked for its lack of performance and asked to try again to see if it could improve. Each child also was asked how long he thought it would take him to do the test this time. In all cases, the variable involved was the discrepancy score between the performance time required for a given trial and the level of aspiration the child set for the next trial. The study concerned itself both with the average discrepancy and with the variability of these discrepancies for each of the groups over a set of twenty trials.

The results showed the success group capable of keeping its aim rather close to target; throughout the experiment they maintained a relatively small positive discrepancy. The failure group, on the other hand, showed a significantly larger discrepancy score between their level of aspiration and their previous performance, and a greater variability in this discrepancy. By comparison with the success group, the failure group scattered its estimates widely in both directions from the performance they might logically have expected to achieve from their previous performance. Their reactions to the frustration situation apparently followed several different patterns: in some cases, it seems as if apathetic behavior developed, perhaps due to the subjects having reached the frustration level, with the result that they then seemed to continue to tolerate a slight positive discrepancy. Some, on the other hand, continued to strive for the improbable, and still others appeared to lose perspective, vacillating from a realistic estimate to estimates that were either unrealistically high or unrealistically low.

Understanding the failure reaction may involve a number of hypotheses. There is obviously no simple formulation that will describe completely the complicated state of affairs; it may well differ from subject to subject and from situation to situation depending on such things as the nature of the individual's self-concept and his ego-involvement in the task. It is possible that some failure subjects felt that their willingness to try would be rewarded—at least for effort, if not for success; they perhaps considered the statement of their goal as a goal in itself. Others may have striven for achievement below what they could do as a means of attaining "success" by doing better than expected. It is also possible that certain subjects, faced with an unpleasant situation, behaved in a trial-and-error fashion; they may even have developed a certain degree of anxiety which may have caused them to lose perspective and to behave in a rather erratic fashion. Others may have continued to pursue impossible goals apparently in a desperate attempt to maintain their self-concept of personal adequacy.

The study has very definite implications for education and should be considered in connection with the self-concept, motivation, and other aspects of the dynamics of student behavior and classroom achievement. It has obvious bearing, for in-

stance, on the matter of grading in the school and on other forms of reward. If we are to accept the conclusions and implications of this experiment, it would seem that success tends to lead the individual to set appropriate goals in line with his abilities and, therefore, success leads to further success. Continued failure, on the other hand, leads him to set goals that are unrealistic and, thereby, to deprive himself of the reward and satisfaction which only true achievement can provide. The results would also mean that the teacher, if he is to promote the development of a positive self-concept and a realistic level of aspiration on the part of his students—and thus place achievement on a self-perpetuating basis—must provide them with individualized—and attainable—goals.

REFERENCE

SEARS, PAULINE S. "Levels of Aspiration in Academically Successful and Unsuccessful Children," *Journal of Abnormal and Social Psychology,* 35 (October, 1940): 498–536.

STUDIES OF UNRELIABILITY IN GRADING

DANIEL STARCH
and EDWARD C. ELLIOTT

Of major importance to education, in view of the prominent position occupied by academic testing in the operation of the school, are the well-known studies of the unreliability of grading essay exams conducted by Starch and Elliott just prior to World War I.

The first study concerned the unreliability of grading English examinations. Two examinations in first-year high-school English were duplicated in their original form, and copies were sent to two hundred high schools in the North Central Association with a request that the principal teacher in first-year English grade the two papers according to the practices and standards of the school. Of the one hundred fifty-two papers returned, ten had to be rejected because of wide differences in the grading standards of the schools involved, leaving a total of one hundred forty-two usable grades. It was further necessary to weight the grades of the schools using 70 as a passing grade to calibrate the grades given to the level of the schools using 75 as a passing grade. The results showed startling discrepancies

—up to some 35–40 points in some cases. Not only were the papers passed by some graders and failed by others, but the order of quality of the two papers was reversed in many instances.

The same two papers were graded by eighty-six university students taking a course on the teaching of English (very few having had teaching experience) ; except for their grading somewhat more leniently, the students gave approximately the same distribution of grades as the teachers. The papers were also graded by a class of superintendents, principals, and teachers taking a course in educational measurements; they also gave a distribution of grades essentially similar to that of the grades given by the teachers.

The wide discrepancy in the grading of these papers led some people to question whether the unreliability was relatively peculiar to grading in the area of English. To test this hypothesis, Starch and Elliott investigated unreliability in grading in the area of geometry, where greater objectivity and accuracy might have been expected. A geometry paper was sent to one hundred eighty high schools in the North Central Association, again with the request that the principal teacher in mathematics grade it according to the practices and standards of the school. A total of one hundred twenty-eight usable returns were obtained. Even greater deviations in grading were obtained than in the English papers; even after adjustment for differences between the passing standards of the various schools involved, the grades ranged from 28 to 92.

In a third investigation, ten papers in freshman English were graded independently by ten instructors of the various sections of college-freshman English. All instructors had given the same final examination, and each had already graded the papers of his own sections. It was found, among other things, that two instructors graded much lower and two instructors graded much higher than the average. Even when the grades were weighted to overcome the lack of uniformity in leniency, sizable differences still existed, especially with respect to two papers which were graded from 44 to 81 for one and from 20 to 65 for the other. Another phase of the study checked the extent to which an instructor agreed with his own grade when he regraded the same paper. An average difference of over four

points was noted. This difference was also found in other subjects, such as language and science.

These studies made a considerable contribution to educational practice in pointing out the unreliability in the grading of the essay examination. They were probably instrumental in the relative wane of the essay examination and the rise of the objective type test. Starch interprets his results as supporting coarser grading—for example, in units of five rather than in units of one or even letter grades, with or without the plus and minus. He was probably also influential in the shift from numerical to letter grades.

REFERENCES

STARCH, DANIEL. "Reliability of Grading Work in Mathematics," *School Review;* 21 (April, 1913): 254–9.

———. "Reliability and Distribution of Grades," *Science* 38 (October 31, 1913): 630–6.

STARCH, DANIEL and ELLIOTT, EDWARD C. "Reliability of the Grades of High School Work in English," *School Review;* 20 (September, 1912): 442–57.

GENETIC STUDIES OF GENIUS

LEWIS M. TERMAN

One of the most significant studies relating to education is the comprehensive investigation of gifted children conducted by Terman and his co-workers under a grant from the Commonwealth Fund and aid from Stanford University. It is published in five volumes, each describing one stage of the investigation. Volume 1, probably the most widely known, analyzes the intellectual and personality traits of 1000 gifted children in central California in the early 1920's. Volumes 3, 4, and 5 represent a follow-up of the same group after five, twenty-five and thirty-five years, respectively. Volume 2 deals with the childhood and youth of the "geniuses" of history and attempts to set a minimum estimate of their likely intellectual level.

The major sample for the first study consisted of 648 gifted children in grades one through eight. A second group of 356 children of the same age living outside the main areas of the study were also included, but less data were collected about them. A third group consisted of 378 high-school students. In addition, a small number of cases were also selected because

of outstanding status in such areas as music and art. A fifth group of 800 non-selected students from the same schools were used as control. On each of the cases in the major sample, some hundred pages of data were gathered, sixty-five pages of which concerned test and measurement data, including two intelligence tests; one education achievement test; tests of general information in the areas of science, history, literature, and the arts; tests of interest in and knowledge of sports and games; reading records, and so on. The additional thirty-five pages concerned questionnaire data regarding home background, school background, medical history, rating of the home and the neighborhood, and so on. In the early phase of the study the minimum IQ required for inclusion in the study was 140; this was later dropped to 132.

The investigation suggested that gifted children, as a group, were superior in all desirable traits; this evidence refuted the belief prevalent at the time that intellectual precocity was generally accompanied by inferiority in non-intellectual areas. Terman's data showed that the gifted group displayed physical superiority, acceleration in school, interest in school subjects (particularly those of an abstract nature), versatility, breadth of reading interest, early maturation, and decisive superiority in such character and personality traits as self-confidence, persistence, and strength of character. They also surpassed normal children in honesty and other moral traits. They showed no lack of interest or ability in sports.

The gifted displayed both hereditary and environmental advantages. The number of eminent relatives among them was impressive: one quarter had relatives in the Hall of Fame. They also came from homes superior in socio-economic status; 81 percent of the parents were professional and semi-professional; 18 percent were skilled and semi-skilled. Most of the children had at least one parent who was a college graduate. Among the interesting sidelights of the study was the fact that the sample included an excess of boys and of firstborn. It also was noted that the investigators could identify the gifted child in a classroom more accurately by selecting the youngest member of the class than could the teacher on the basis of judgment.

The second volume of the series deals with the study of 301 outstanding men in history with a view to discovering the mini-

mum level of mental endowment that they must have possessed in order to have accomplished what they did. A case study was compiled for each of the subjects; it was found, for example, that many of them had been able to read at the age of three or had studied Greek and Latin at a pre-school age. The geniuses of history gave evidence both of superior hereditary and of environmental advantages; they also displayed the usual characteristics of the gifted.

The third volume was a five-year follow-up and included 97 percent of the original group of gifted children. The study duplicated essentially the same measurements. It was found, for instance, that the average IQ of the group had dropped slightly, as might have been expected on the basis of the phenomenon of regression toward the mean.

The fourth volume reports the findings of the twenty-five-year follow-up, in which data were again collected through information blanks, interest and personality inventories, and other means. It was found, for instance, that the offspring of the gifted sample had an average IQ of approximately 127, which, again, is in line with the concept of regression toward the mean. Among the findings of this twenty-five-year follow-up, there was evidence of greater marital adjustment among the gifted than for the general population, a lower death rate, a lower incidence of delinquency, a better record of employment, a higher level of professional accomplishment, and a continuation of such personality characteristics as a sense of humor, cheerfulness, optimism, will-power and perseverance, desire to excel, and self-confidence. They gave every indication that giftedness in youth is a fairly good indication of similar giftedness throughout life. On the other hand, a number of these gifted children did not achieve in keeping with their potentialities, but these could not be distinguished from the more successful with respect to intellectual status. Whatever differences were involved appeared to center around such personality characteristics as drive and perseverance. The fifth volume presents a thirty-five-year follow-up of the group and suggests a continuation of the same life of success and outstanding achievements.

Terman's study is an especially good example of a carefully planned and executed longitudinal investigation. His success in

obtaining co-operation and maintaining contact with his subjects is in large part a reflection of the intellectual and cultural level of the subjects involved. The study is obviously a classic in the field; its contribution to the understanding of the gifted child is becoming of greater importance with the present emphasis upon the education of the gifted. Not only has it answered a number of questions concerning the gifted, but it has also set the pattern for further investigation in this area.

REFERENCES

Burks, Barbara S., *et al. Genetic Studies of Genius 3. The Promise of Youth.* Stanford: Stanford University Press, 1930.

Cox, Catherine M. *Genetic Studies of Genius 2. The Early Mental Traits of 300 Geniuses.* Stanford: Stanford University Press, 1926.

Terman, Lewis M. *Genetic Studies of Genius 5. The Gifted Group at Mid-Life.* Stanford: Stanford University Press, 1959.

Terman, Lewis M. and Oden, Melita H. *Genetic Studies of Genius 4. The Gifted Child Grows Up.* Stanford: Stanford University Press, 1947.

Terman, Lewis M., *et al. Genetic Studies of Genius 1. Mental and Physical Traits of 1000 Gifted Children.* Stanford: Stanford University Press, 1925.

THE DISCIPLINARY VALUE OF HIGH-SCHOOL STUDIES

Edward L. Thorndike

Thorndike's early work in educational psychology, particularly in transfer of training, led him to question the transfer value of such classical subjects as Latin, the sciences, and mathematics. He undertook a study to determine the relative value of the various high-school subjects in the improvement of reasoning ability. The study was based on the comparison of the gains in reasoning made by students enrolled in the various high-school curricula. It involved the testing of over 8000 pupils in grades nine through eleven with Form A of the I.E.R. Test of Selective Relational Thinking and the I.E.R. Test of Generalization and Organization in the fall of 1922, and again in the spring of 1923 with Form B of the same tests. The study was repeated under Thorndike's direction by Broyler and Woodyard, the sample in this case being approximately 5000. In both studies, the gain in reasoning ability of the group enrolled in business, drawing, English, history,

music, shop, and Spanish was used as a point of reference, and the gains from the other curricula were measured as an index of departure from this benchmark. Unfortunately, the two studies did not show any great degree of consistency in the ranking of the subject areas from the standpoint of the gain in reasoning ability which they promoted. Combining the results of the two studies, the subject areas came out in the following order: (1) algebra, geometry, trigonometry; (2) civics, economics, psychology, sociology; (3) chemistry, physics, general sciences; (4) arithmetic, bookkeeping; (5) physical training; (6) Latin and French; business, drawing, English, history, music, shop, Spanish; (7) cooking, sewing, stenography; (8) biology, physiology, agriculture; and (9) dramatic art. However, the differences in all cases were relatively small. It was Thorndike's conclusion that "the expectation of any large difference in general improvement of the mind from one study rather than another seemed doomed to disappointment." He pointed out further that the balance in favor of any study is certainly not large; disciplinary values may be real and deserve weight in planning the curriculum, but the weight should be reasonable.

A more adequate study of the same problem is that of Wesman who overcame one of the basic weaknesses of the previous studies—that is, the fact that no distinction had been made between being merely enrolled in a given curriculum and mastering its content. Wesman's study duplicated the other two, with the added feature of measuring the degree of achievement in the subjects being compared. His results agree rather closely with those of the two previous studies and, in general, confirm the conclusion that the transfer or disciplinary value of the various academic curricula in the improvement of reasoning ability is not appreciably different from subject to subject. He concluded that "in general, the study fails to reveal superior transfer to intelligence for any one of the achievement areas measured and indicates the desirability of direct training in mental processes rather than dependence on transfer from school subjects."

These studies constitute an important landmark from the standpoint of curriculum construction and mark the beginning of the de-emphasis of the classic subjects on the strength of their disciplinary power. It is the general consensus today that there

is no superior subject matter for transfer; rather there are only superior learning experiences, and transfer is more a function of the way learning takes place than of the subject matter involved. It should also be noted that in all three studies a substantial relationship was noted between the extent of transfer and the intellectual status of the individual.

REFERENCES

BROYLER, CECIL R., *et al.* "A Second Study of Mental Discipline in High School Studies," *Journal of Educational Psychology,* 18 (September, 1927): 377–404.

THORNDIKE, EDWARD L. "Mental Discipline in High School Studies," *Journal of Educational Psychology,* 15 (January and February, 1924): 1–22, 83–98.

WESMAN, ALEXANDER G. "A Study of Transfer of Training from High School Subjects to Intelligence," *Journal of Educational Psychology,* 39 (December, 1945): 254–64.

THE LAWS OF LEARNING

EDWARD L. THORNDIKE

Probably no other research studies have had more effect on American education than have Thorndike's studies in learning. They mark the beginning of the modern experimental attack on the problem of learning and of the modern development of the theory of learning. Although his experiments were conducted on animal subjects, they had tremendous impact on educational practice in the first decades of this century, and are valuable even today.

His investigations comprised a number of experimental studies with various animals and were designed to determine the factors governing the process of learning. In some of his early studies, he had fishes learn to go through a small hole in a dividing plate in order to get away from the sunlight into a shady spot. Motivated by their aversion to strong light, the fish gradually learned to find the hole with greater and greater success. In his experiments with chicks, Thorndike had them find their way through mazes in order to get food. In all cases, the animals showed a progressive, though erratic, decline in the time they needed to attain a solution.

Probably his best known experiments are those in which a hungry cat was placed in a box from which it could free itself by pulling a string, pressing a lever, or turning a button.

After trying to squeeze through the slats of the box or clawing at the door, the cat finally hit on the escape device—generally by accident—and was able to get the food. As it was placed in the box in later trials, its activities became progressively more directly oriented toward the escape mechanism until, after a number of trials, its escape was relatively automatic. In all cases, however, learning proceeded largely through the gradual elimination of error from a trial-and-error approach; the display of insight, as Thorndike saw it, was minimal, if not completely lacking.

Thorndike's experiments have been subjected to some degree of criticism; it is felt, for instance, that his interpretation of learning is too mechanistic, that it does not give enough credit to the learner for his reasoning ability. It is possible, for example, that Thorndike's cats were young and relatively untamed, causing them to be more excitable and erratic than older and tamer animals. An even more pertinent objection is that the task which he imposed on the cats was relatively devoid of means-ends relationships into which they could develop insight. In a similar experiment with dogs, Thorndike found them to exemplify essentially the same learning pattern as the cats, though their progress tended to be somewhat smoother, perhaps because of their greater intelligence. On the other hand, the dogs did not learn as fast as monkeys and racoons, possibly because of the greater curiosity of the latter and their greater ability to manipulate mechanical devices because of the construction of their paws.

In a similar study with monkeys, in which he used such signals as giving them food when the experimenter had the food in his left hand but not when it was in his right hand, Thorndike noted that the monkeys could discriminate between certain signals but not others. He felt that the reasoning ability of animals was relatively limited, and that trial-and-error rather than insight was the dominant factor in learning. These findings have been partially contradicted by other investigations, at least from the standpoint of emphasis. The issue seems to revolve around the definition of reasoning; apparently apes and monkeys are able to solve problems which demand a fairly high degree of abstraction, or ability to discover relationships, as well as memory, but whether that constitutes rea-

soning is a matter of definition. Thorndike was unduly pessimistic in this connection and conceived of learning as being relatively mechanical. Köhler, on the other hand, was able to demonstrate a number of instances of reasoning or "insight" in chimpanzees, such as piling boxes one on top of another in order to reach a banana on the ceiling of the cage or putting two sticks together in order to reach a banana placed outside the cage.

Thorndike's principal contributions to psychological science are his two major laws of learning: *the law of exercise* and *the law of effect*. The *law of exercise* stated that the more frequently a neural connection is used, the stronger it becomes. Its counterpart, the *law of disuse,* postulated that when a connection is not used it becomes weakened. It is best considered as a factor in memory and forgetting rather than of learning *per se*. The *law of exercise* (use) has been criticized by a number of psychologists on the grounds that even often-repeated activities are frequently not learned, a man may look at his watch twenty-five times a day and still not know whether the numbers are Arabic or Roman. The concept of exercise, as we see it today, is the principle which describes how learning is acquired under certain circumstances rather than as a cause of learning; it is considered a necessary but not sufficient condition for learning to take place.

Thorndike's *law of effect,* originally had two components, namely (1) S-R (stimulus-response) bonds followed by satisfying after-effects tend to be strengthened, and (2) S-R bonds followed by annoying after-effects tend to be weakened. (The latter component was discarded.) The *law of effect* is, generally speaking, the most important law in psychology of learning, relating as it does to the concept of motivation. It has, of course, not received complete endorsement; some psychologists prefer the more general concept of reinforcement which is, in a sense, not radically different from the concept of effect.

Nevertheless, regardless of its theoretical validity, the law of effect, in its broad sense, has unquestioned operational usefulness and is generally accepted by practitioners in the field. The law of exercise, on the other hand, is somewhat less acceptable. At one time, it served as the basis for the justification of drill as the basis of instruction, a practice which has more or

less been superseded by the present emphasis on meaningfulness. It is now realized that only meaningful practice leads to improvement. Both of these laws have been subjected to considerable discussion in the psychological literature, and no clearcut evaluation of their status can be given in the short space here. In general, the major objections to Thorndike's laws of learning center around the relatively mechanistic interpretation of learning which they imply. A clear understanding of their nature would call for a thorough background in the psychology of learning.

REFERENCES

HILGARD, ERNEST R. *Theories of Learning.* 2nd ed.; New York: Appleton-Century-Crofts, 1956.

KÖHLER, WOLFGANG. *The Mentality of Apes.* New York: Harcourt-Brace, 1928.

MARX, MELVIN H. "Spread of Effect: A Critical Review," *Genetic Psychological Monographs,* 53 (February, 1956): 119–86.

POSTMAN, LEO J. "The History and Present Status of the Law of Effect," *Psychological Bulletin,* 44 (November, 1947): 489–563.

THORNDIKE, EDWARD L. *Animal Intelligence.* New York: Macmillan, 1911.

———. *Human Learning.* New York: Appleton, 1931.

WOODWORTH, ROBERT S. *Contemporary Schools of Psychology.* rev. ed.; New York: Ronald Press, 1948.

THE TEACHER'S WORD BOOK

EDWARD L. THORNDIKE

Of interest to teachers in the elementary school are Thorndike's three vocabulary studies in which he identified the 10,000, 20,000, and 30,000 most widely used words. Although some work in this direction had been done prior to his, and although more adequate studies have been conducted since, Thorndike's studies still remain among the better known. The magnitude of the task alone is frightening. The first word list (10,000 words), published in 1921, was selected from a count of "about 625,000 words from the literature of children; about 3,000,000 words from the Bible and English Classics; and about 3,000,000 words from elementary-school textbooks; about 50,000 words from books about cooking, sewing, farming, the trades and the like; about 90,000 words from the daily newspapers; and about 500,000 words from correspondence."

The words are classified into 1000, 2000, 3000, . . . levels according to frequency of occurrence, with a further breakdown of the first 1000 words into the various 100 levels, and the next 4000 in 500 levels of frequency of use. The words are also rated on "range" or extent of use in the forty-one sources consulted in the derivation of the list. His word lists of 20,000 1932) and of 30,000 words (1944) are essentially a continuation of his first study. Again, the words are rated as to range and frequency of use.

Thorndike felt the word lists would be of benefit to the teacher in judging the importance and difficulty of a given word as the basis for deciding on the emphasis to be placed on it at a given grade level. He pointed out emphatically that they were not to be construed as spelling lists. He acknowledged that the lists were not perfect measures of the relative importance of the word listed. A word may have personal interest and value for a student and yet not be of common currency in the world's readings. He also pointed out that tens of thousands of hours of further counting would have been necessary to measure the frequency of use of words with exactness. He acknowledged the possibility that some one thousand words might be more deserving of inclusion in the lists, and that the order of the lists might be changed somewhat. He had particular trouble with the correct placement of proper names, new words, and abbreviations. It would also follow that the list would become out-of-date with the inclusion of new words and the deletion of older words. It would also be true that regional differences—for example, rural-urban differences—exist, which make such a list somewhat short of perfect as a guide. Nevertheless, the lists, in addition to promoting further studies in this area, have had definite influence on the concept of controlled vocabulary load, particularly in the primary grades and, of course, more recently in the field of readability.

REFERENCES

DALE, EDGAR. *Bibliography of Vocabulary Studies.* Columbus: Bureau of Educational Research, Ohio State University, 1949. (Also Dale, Edgar, and Reichert, Donald, 1957.)

LINDSAY, EDWARD Y. *An Etymological Study of Ten Thousand Words* in *Thorndike's Teacher Word Book.* Indiana University Studies, 12: No. 65. Bloomington: Indiana University, 1925.

Thorndike, Edward L. *The Teacher's Word Book.* New York: Teachers College, Columbia University, 1921.
———. *The Teacher's Word Book of 20,000 Words.* New York: Teachers College, Columbia University, 1932.
———. "Word Knowledge in the Elementary School," *T. C. Record,* 222 (September, 1921) : 334–70.
———. *The Teacher's Word Book of 30,000 Words.* New York: Teachers College, Columbia University, 1944.

FACTORIAL STUDIES OF THE MIND

Louis L. Thurstone
and Thelma G. Thurstone

Particularly significant to vocational counsellors are the investigations of the nature of intelligence conducted by the Thurstones, in which they attempted to identify the basic components of the mind. These and later studies have isolated a number of mental factors or abilities, primary among which, of course, is the verbal factor (V). Other factors commonly accepted include the numerical factor (N), the word fluency factor (W), the spatial factor (S), the memory factor (M), the reasoning factor (R), and the perceptual factor (P).

The major study involved in the identification of the primary mental abilities is a study conducted in the thirties, involving sixty tests, many of which were devised specifically for the study. In addition, data on chronological age, mental age, and sex were added for a total of 63 variables to be processed through factor analysis. The subjects were 710 out of 1154 eighth-grade children for whom complete records were available. The factors were rotated by means of the common centroid method of factor analysis to provide seven identifiable factors (N,W,S,V,M,R, and P), and three relatively indeterminate factors. The results were checked in a second factorial study involving 437 cases, the data for which were again processed by factor analysis to yield the same factors.

The Thurstones have made valuable contributions in a large number of vital psychological areas, particularly the measurement of attitudes, interest, and intelligence. They also have made equally worthwhile contributions in the development of rating scales, scale analysis, and factor analysis. Most of the work with which we are concerned here is based on factor analysis, and an appraisal of their studies requires an under-

standing of the strength and limitations of factor analysis, some of which have been mentioned in Chapter 11. Nevertheless, their contributions constitute a valuable addition to the field of psychometrics. From a theoretical point of view, their position regarding intelligence is in apparent conflict with that of Spearman, though the difference is one of orientation and emphasis rather than of outright disagreement. The Thurstone approach is particularly appropriate in vocational guidance which is based, in large measure, on the differential adequacy of the counselee's aptitudes or primary mental abilities.

REFERENCES

THURSTONE, LOUIS L. *Primary Mental Abilities.* Chicago; University of Chicago Press, 1938.

THURSTONE, LOUIS L. and THURSTONE, THELMA G. *Factorial Studies of Intelligence.* Psychometric Monograph No. 2. Chicago: University of Chicago Press, 1941.

TEACHER'S ATTITUDES TOWARD CHILDREN'S BEHAVIORAL PROBLEMS

E. K. WICKMAN

This widely quoted study was conducted in 1928 under the auspices of the Commonwealth Fund. A total of 511 teachers and 30 clinicians rated fifty behavior problems on their relative severity. The results revealed a substantial reversal in the ratings of teachers and clinicians: of the twelve behavior problems rated most severe by teachers, two were rated among the least severe by the clinicians; and of the twelve behavior problems rated least severe by teachers, three were rated among the most severe by the clinicians. In general, it might be said that teachers considered shyness, sensitivity, unsociability, fearfulness, dreaminess, and other purely "personal" problems which did not interfere with the teacher's immediate purpose among the least serious, while the clinicians rated these factors among the most serious. On the other hand, teachers rated behavior problems relating to sex, dishonesty, and disobedience as much more serious than did the clinicians. More simply, teachers placed emphasis on anti-social tendencies (defiance of authority and violation of rules); clinicians, on the contrary, saw greatest danger in unsocial tendencies (shyness and sensitivity). The study revealed, for example, a tendency for teachers to

counterattack the "attacking" type of problem behavior and to indulge habits of withdrawal and dependency—thus aggravating both unhealthy conditions.

The study is often quoted as evidence of the fact that teachers do not understand what constitutes a serious mental health problem. This interpretation is relatively gratuitous, revealing some lack of understanding of the nature of the study. Wickman points to differences in professional interest; the clinician is interested in the social and emotional adjustment of the child while the teacher is interested in his educational accomplishment.

Wickman points very clearly to differences in the directions given to the two groups: "(1) the directions to teachers for rating were phrased in such a way as to secure responses to the *present* problem and the question of the significance of the present behavior disorder upon the *future* development of the child, though possibly unavoidably implied, was not definitely raised. The task set was to rate the degree of maladjustment represented by the immediate problem. (2) Care was also taken to establish in the teachers a mental set for responding to the "seriousness of," the amount of "difficulty produced by" the particular type of troublesome behavior. The assumption was that the degree to which teachers found a certain trait serious, difficult, or undesirable represents the amount of attention they directed to the problem and the effort exerted toward its modification. (3) Then, too, in order to elicit the first, unrationalized reactions, the teachers were instructed to rate as rapidly as possible and a time limit was imposed for completing the ratings." The clinicians, on the other hand, instead of evaluating the present problem, rated its significance with respect to its future effects in limiting a child's happiness, success, and general welfare after leaving school and on entering adult life. The differences in ratings between teachers and clinicians become more understandable in the light of the differences in the directions which they received.

It should be noted that more recent studies incorporating equivalence in the directions given both to the clinicians and to the teachers have revealed much closer agreement in their attitudes as to the severity of children's behavior problems. Schrupp and Gjerde, in a repetition of Wickman's study using the same

set of directions for both groups, found a correlation of 0.56 in the ratings of the two groups in contrast to the correlation of −0.04 found by Wickman. Whether one needs to be disturbed over the discrepancy in outlook that still exists between teachers and clinicians is a matter of opinion. There may be a need for a further shift of teachers from concern over breaches of classroom decorum to a more objective consideration of behavior from the standpoint of the long-term development of the whole child. On the other hand, it is questionable whether there should ever be complete agreement in view of their different purposes and the different settings in which they operate.

REFERENCES

SCHRUPP, MANFRED H. and GJERDE, CLAYTON M., "Teacher Growth in Attitudes towards Behavior Problems of Children," *Journal of Educational Psychology,* 44 (April, 1953): 204–14.

WICKMAN, E. K. *Children's Behavior and Teacher's Attitudes.* New York: Commonwealth Fund, 1928.

Appendix

The Thesis and Dissertation

The research report with which most people studying this text are concerned is the master's thesis or the doctoral dissertation. It, therefore, seems advisable to devote a section to this important project, especially since, though such a project should generally be "the accomplishment of a lifetime," the end-result is too frequently a disappointment.

Unfortunately, no magic formula can be given that will ensure an adequate product, and, though this section will attempt to present a few ideas on the various aspects of conducting and reporting research, these ideas are simply suggestions. Generally, the giving of specific rules is incompatible with the whole process of research which, to be fruitful, must remain flexible. Not only is a good research study not to be equated with step-by-step directions, but further, anyone who needs such help probably should not be doing research in the first place.

THE RESEARCH PROPOSAL

Research starts with the identification of a problem. This is the first step in the sequence; it is also of prime importance, for probably no aspect of the study has a greater bearing on the success of the overall venture than the wise choice of a problem. As we have seen in Chapter 4, it is both the prerogative and the responsibility of the graduate student to identify a suitable topic; devise a suitable plan of attack; collect, process, and interpret the required data; and finally to write the report. It is

here that the student will establish his claim to status as a leader in the profession.

The early stages of the selection of a topic is generally a matter of trial-and-error as the student traces one lead after another, dropping an idea here, gradually developing another, until finally a topic emerges. An important part of this search is discussing ideas with others, whose reactions invariably lead to a refinement of the proposed study. Research seminars conducted for students working on their dissertations can be of great help in this connection. There is probably no better way of having the student clarify his thinking on every phase of his proposed investigation than requiring him to defend his proposal before such a seminar, prior to submission for approval by his committee. A key person in the process of choosing a topic is the student's advisor, who, because of his familiarity with the field, can generally save the student much aimless wandering and exploration of blind alleys. Major professors in the field are also valuable sources of help in formulating a tentative proposal for approval by the advisor and, eventually, by the committee.

On the other hand, the student seeking faculty help should come with ideas to discuss and with tentative topics clearly formulated. These ideas generally should be presented in writing as evidence that he has done some thinking about his problem. It is rather difficult to give constructive advice to the student who "wants to write a thesis . . . ," who isn't sure of the area, "Administration, I guess . . . —Maybe something in how to deal with personnel." Before he can be helped—short of being handed a topic ready-made—he needs to get his ideas more clearly defined.

The proposal submitted for committee approval should be sufficiently detailed and clear that it is an actual blueprint for the study to follow. Generally, the proposal should present the general nature and the present status of the problem, the theoretical and empirical framework within which it exists, the hypotheses to be tested, its significance and likely contribution, its feasibility, the method of attack (including the proposed analysis of the data), and so forth. Although this is a proposal and not the final draft of the thesis or dissertation, it must give evidence of careful planning and anticipation of

problems. Deviations from the original plans may have to be made, but meticulous formulation of the proposal will keep such modification to an absolute minimum.

WRITING THE REPORT

The specific arrangement of the thesis or dissertation from the standpoint of such details as chapter organization varies from topic to topic. Thus, a short master's thesis may be organized in three chapters: 1. The Problem; 2. The Design and Results of the Study; 3. The Summary and Conclusions. A doctoral dissertation, on the other hand, may have as many chapters as a good-sized book. Separate chapters may be devoted to the review of the literature, to the findings, and to the interpretation of the data, for example.

The length of the report also varies. Historical and survey studies, for instance, tend to be longer than experimental studies. The criterion is not the number of pages, however, but the adequacy of the scope of the problem and its treatment. The author is reminded of the anecdote of the magazine editor who received a call from a neophyte writer inquiring about the length of the average novel. When the editor answered that the usual novel ran from 75,000 to 95,000 words, the young voice at the other end of the line exclaimed: "Well then, I've finished!"[1]

The general format also varies somewhat with the topic, its nature, its scope, and its complexity, as well as with the individual preferences of the writer. Generally, however, the research report divides itself into three major parts: 1. the preliminary section which includes the title page, the acknowledgments, the table of contents, and the list of tables; 2. the report itself, which is divided into the introduction, the statement and delineation of the problem, the development of the data, and the conclusions; and 3. the bibliography and other supplementary material.

The Problem

The first section of the report must present the problem, its nature, its scope, and its significance. Usually this section be-

[1] Jacques Barzun, and Henry F. Graff, *The Modern Researcher* (New York: Harcourt, Brace, 1957), p. 19.

gins with a general orientation to the problem area and leads directly into a statement of the problem to be investigated. This section should be appealing and challenging, and generally it is difficult to write it well.

The statement of the problem is crucial since it delineates specifically what is to be investigated and, thus, what is relevant, what is irrelevant, and what constitutes an effective approach to its solution. The problem should be stated directly and, wherever possible, should be translated into a specific hypothesis to be investigated. The statement must distinguish clearly between what the study will investigate and what it will exclude from consideration. It should also leave no doubt about the assumptions being made. The statement should come early: it is frustrating to read page after page of discussion before finding out what the problem is that will make such discussion relevant. Special terms or special meanings given to common words should be clearly defined in the interest of clarity.

A section which is frequently under-emphasized is that on the significance of the problem—that is, the justification of the proposed study and its implications for educational practice. While to the student the study may be all-important, the reader may not be enlightened about its significance and the possible contributions the study may make. This should be included in the proposal and, of course, should be integrated with the implications of the study discussed in the final chapter.

The Review of the Literature

As we have seen, the review of the literature is essential to the development of the problem and to the derivation of an effective approach to its solution. Not only must this section be thorough and exhaustive, but it must also be organized with subheadings which will structure the literature with respect to the specific aspects of the problem. The review of the various sources should be analytical rather than merely cumulative—that is, the studies should be evaluated on the basis of adequacy and relevance rather than simply listed. Furthermore, the various sources must be integrated and synthesized to give the reader a clear picture of the status of the problem as background for placing the present investigation

in perspective. It might be said that any review of the literature that is left hanging—that does not eventuate in a clearcut generalization—is incomplete. Often, the literature is so extensive that the student must be selective about what he includes; while this involves certain risks, the writer should be in a position to make a judgment on whether listing one study after another on a given point is likely to clarify or to becloud the issue.

Generally, students do a rather thorough job of reviewing the empirical literature. Frequently, however, they do not present an adequate conceptual framework within which their problem exists, nor do they relate their findings to their theoretical implications. For example, in investigating the personality characteristics of non-readers, there is apparently an assumption of reciprocal interaction of non-reading and personality maladjustment. On the other hand, an investigation of good readers might be approached from the hypothesis that an exceptionally good reader may be either a well-adjusted individual or a maladjusted individual who compensates through overachievement in reading. These hypotheses are, of course, illustrative only; the point is that it is best not simply to investigate the relationship between reading and personality without presenting the hypothetical conception from which the study might have originated.

The Design

A study cannot be evaluated unless its procedures are reported in sufficient detail to make such an evaluation possible. The section on the design should be particularly clear and precise to allow the reader to grasp exactly what was done and, in the event of a need for verification or refutation, to permit its exact replication. Furthermore, since a primary purpose of reporting research is to permit a more effective attack on a given problem, the procedures used in a study should be reported, even to the point of recording the more important blind alleys that were tried and abandoned.

The specific aspects of the design that need to be emphasized vary with the nature of the study. In a survey study, for example, it is generally essential to describe the locale in which the study is conducted, for the findings of a survey can-

not be interpreted apart from such a consideration. In a survey of the reactions of teachers to the twelve-month school year, it would be essential to know whether the school buildings are air-conditioned, whether the community is rural or urban, and so on, just as a survey of the reactions of teachers toward merit pay would have meaning only on the basis of the present morale of the teachers, the facilities for adequate teacher evaluation, as well as the specific plan of merit pay proposed. On the other hand, the locale would be less important in an experimental study of the effects of increased emphasis on phonics in reading in the primary grades. Here, however, the specific exercises incorporated into the "experimental" method in contrast to those of the "control" method would have to be described at length. Sample lessons and a detailed list of guidelines and principles for the guidance of the participating teachers might also be included. Similarly, the specific methods used in selecting a random sample are most important in a normative-survey study, while the establishment of the equivalence of the experimental and control groups is the point to emphasize in an experimental design.

When instruments are used, they should be described from the standpoint of their validity in the present case, their reliability, and their standardization. The report of a questionnaire study should indicate the specific steps taken to devise and to improve the questionnaire and should provide evidence of the adequacy of the final product. Copies of all but well-known instruments should be included in the appendices. In a questionnaire study, copies of the cover letters and the follow-up attempts should also be included, since they have direct bearing on what was done and on the results achieved. When descriptions of this kind are particularly lengthy, a few summary statements should be made in the text of the report and a more complete description included in the appendix.

Collection and Interpretation of Data

No study can be better than the data on which it is based and the interpretation which they are given. The adequacy of the data relates not only to the adequacy of the research design but, even more directly, to the adequacy of the instruments used. Skill in the choice and the use of research instru-

ments is, therefore, crucial to the success of the study and to the validity of its results and conclusions. Important here is the proper use of the instrument as a measuring device and, since the validity of an instrument and of the data which it yields is a concept specific to the circumstance in which it is used, the proper interpretation of the results in the light of the problem. In an experimental study, if the instrument is not equally "fair" to the methods being compared, only misinformation about the relative effectiveness of the methods being compared can be obtained. In view of the very crucial role which instruments of measurements play in the conduct of research, anyone without background in the theory and practice of tests and measurements is bound to be very much restricted as a research worker.

Summary and Conclusions

The final section of the report proper consists of a review of the significant aspects of the whole study structured so that it leads directly to the conclusions. This summary is important in that it places the whole study in perspective. It must be carefully written, especially since frequently this is the only section of the report that the busy person reads.

No part of a study is more important than any other part, since a defect in any part will automatically affect the whole study. If, however, one part can be singled out as all-important, it is the section which states the conclusions, for this is the section that presents what the study has to contribute to the advancement of education as a science. It is frequently a very difficult section to write, inasmuch as it must be accurate and precise, as well as insightful.

Drawing the conclusions is a matter of clarifying "Just what did the study reveal?" The conclusions can be a simple answer to the question or hypothesis raised in the statement of the problem. In more complex designs and, especially in a survey study where the findings are usually multiple rather than single, the student must orient himself to the significant aspects of his study. Ordinarily, the findings should be organized into a relatively small number of meaningful groupings of ideas, each with a definite significance for the problem under study. Particularly useless—in fact, confusing—from the standpoint

of the conclusions to follow, for example, is the long recitation of "findings" cited pell-mell without regard to their implications or their relative significance.

Once the findings have been organized into a half-dozen clusters, the conclusions should follow directly. If the student has difficulty with his conclusions, it is generally because he is not clear as to his findings. Perhaps he needs to go back and analyze his data further; he may need to establish more clearly the interrelationships among his findings and between his findings and those of other investigators. He may even have to get a clearer perspective of the whole field. (Incidentally, a very common fault in research reports is that of confusing findings with conclusions and vice versa; they are, of course, different, and any student who takes the time should be able to tell them apart.)

It must also be noted that, at this point, the investigator abandons his role as a scientist and becomes a philosopher. This, of course, puts him in a vulnerable position since interpretation always involves an element of subjectivity. On the other hand, this is a responsibility which the investigator has to assume, for certainly he should have a more adequate understanding of his area of investigation than anyone else, and he is obligated to provide an interpretation of the meaning of his findings. He is, for example, in the best position to see the limitations of his study and to point out the need for certain cautions in the interpretation and application of his findings. He should also be capable of pointing out the direction which further research may take, the pitfalls to be avoided, and the best utilization of research talent in the pursuit of the solution. This is an area in which the investigator can make a real contribution to the field. Certainly, he must have learned something from his experience that he can share with future investigators.

The conclusions are the expression of the investigator's personal interpretation of the facts he has uncovered. The object is to establish as clear-cut an answer to the questions posed in the statement of the problem as the data of the present study, analyzed in the light of the situation and of the work of previous investigators, will permit. The investigator must be particularly careful not to go beyond his findings

and project his personal biases into the data. He must be alert to such errors of logic as confusing concomitance for causation or the cause for the effect (reverse causality). He must not disregard contradictory evidence, nor must he fail to recognize the limitations of his study. He must maintain his objectivity and, if faced with negative evidence, admit readily that his hypothesis was in error, rather than make lame excuses about the inadequacy of the study.

A common shortcoming of research reports is the failure of the writer to relate the findings and conclusions of the present study to those of other investigators presented in the review of the literature. It is essential that the conclusions of the study constitute the final word on the subject, incorporating all research on the subject to date (including the present study). The findings and conclusions of the present study must be integrated—and reconciled where necessary—with those of previous investigators. Any limitations which might have influenced his results or those of others should be clearly pointed out. No study is perfect: compromises from an ideal design generally have to be made to conform to the reality of the situation. The important thing is that a clear exposition of the present empirical and theoretical status of the problem be presented with a definite statement of the degree of confidence that can be placed in what appears to be known about it.

MECHANICS OF THE REPORT

Importance of Scholarship

Graduate students sometimes fail to appreciate the importance of having the report radiate the same high level of scholarship that went into the investigation itself. Nothing can detract from a good study more than carelessness in its report. Conversely, in the absence of evidence to the contrary, the display of carelessness and incompetence in the report is of itself sufficient grounds for suspicion of equal carelessness and incompetence in the conduct of the study. What is more, such carelessness is inexcusable in a candidate for a graduate degree.

A major aspect of faulty reporting concerns grammatical usage—that is, failure to adhere to accepted rules of gram-

matical structure—coherent organization, and attention to the many details necessary for producing a first-rate piece of work. Another failure is a lack of precision and effectiveness in expression. Although there appears to be a close relation between effective writing and clear thinking, it is also probably true that clear, concise, and forceful expression does not come, even to the clearest thinker, without the expenditure of considerable time and effort. Regardless of his literary talent, the writer will invariably find that many revisions are necessary to bring the report into acceptable form. Attention to details—major and minor—is the price one must pay for scholarly work, and this insistence on quality is one of the major features that distinguishes the graduate from the undergraduate student.

The writing of the thesis or dissertation is governed by a number of regulations; some merely emphasize the obvious while others are relatively mechanical and arbitrary. In the main, these regulations make good sense, and failure to comply generally invites trouble. Practice and experience have led to the development of a format and an organization which are designed to promote maximum clarity in reporting and maximum effectiveness of use by the reader, who can then devote his whole attention to the content. For example, the statement of the problem comes first because only when the reader knows what the problem is, can he evaluate the relevance of the literature reviewed and the adequacy of the research design. Similarly, the fact that the last chapter contains the study in a nutshell is a boon to the busy reader. This is not a matter of stifling the writer: there is plenty of opportunity to show one's ingenuity within the framework of a uniform format. Indeed, there is no limit to the extent to which the writer can display originality, creativity, and initiative in the design of the study, in the style of presentation, in the choice of vocabulary, and so on. Would that the student display more of it!

Format

The format of the report is generally specified in some detail by each individual school. The sample pages and sugges-

tions provided on pages 490–492, are therefore, simply acceptable models from which deviations to comply with local regulations will probably have to be made. Some graduate schools are more specific than others. Many use a standard style; some have their own manuals of style; others allow the student to use any acceptable style, provided he is consistent in the use of the style he adopts. The differences that do exist tend to concern details rather than major points. For example, there is agreement on the need to include volume number, page numbers, and date of an article in a bibliographic entry; disagreement is frequently found, however, with respect to the order in which they are to be listed.

In general, it may be argued that the style adopted for educational writings should conform to the usual style of writing in the field of education. There might be value, for instance, in adopting a style which is consistent with that of the major publishers of textbooks in education. It should, of course, meet such criteria as completeness, uniformity, convenience, clarity, and appearance. Thus, it might be acceptable to omit the publisher in a footnote notation, since this information is included in the bibliography, but the date should not be omitted, since it may be important to know if the item was written in 1900 or in 1960.

The Title Page. The importance of the wise choice of a title is sometimes overlooked. Theses and dissertations sometimes carry titles of such comprehensiveness that the reader wonders if the writer is attempting to write his whole report on the title page. The title must be sufficiently indicative of the study that it does not mislead the reader. Note, for example, how annoying it is to run down a source that, from its title, seems relevant only to find that its content is on an entirely different subject. There is also the danger of the research worker overlooking an article of importance to his study simply because its title made it appear irrelevant. The general format of a title page is shown in the specimen page. Note, for example, that the title is spaced in inverted pyramid when its length calls for more than one line of writing.

The Acknowledgments. The acknowledgment page frequently appears to clash with the objective and scientific tone

THE RELATIONSHIP OF SOCIO-ECONOMIC STATUS

TO PERFORMANCE ON THE ITEMS OF

THE REVISED STANFORD-BINET

A Thesis Submitted to the Graduate Faculty of

the University of . . .

in Partial Fulfillment of the Requirements

for the Degree of

Master of Education

by

John C. Doe

City, State

Month, Year

TABLE OF CONTENTS

CHAPTER 1

THE PROBLEM AND ITS BACKGROUND

Few scientific problems have been the subject of so much speculation and controversy as have the estimation of the influence of environment on intelligence, and the determination of the degree to which intelligence tests now in common use can be considered valid instruments for the measurement of the mental ability of subjects who deviate widely in one or more respects from the average of the group on which the norms of such tests were derived. Inasmuch as the Revised Stanford-Binet Scale of Intelligence is used extensively to test children from all socio-economic levels, there is a need to investigate the reliance which can be placed on the results of this scale when it is used to test pupils whose socio-economic backgrounds are either extremely favorable or extremely unfavorable.

The Problem

Statement and delimitation of the problem. It was the purpose of this investigation to compare the performance on the items of the Revised Stanford-Binet Scale of Intelligence, Form L, of pupils from homes of low socio-economic status with that of pupils of equal mental age coming from homes of high socio-economic status, with a view to determining the items, and particularly the types of items, if any, on which performance is greatly affected by differences in so-

which a paper of this kind should reflect. Although it is natural for the student to feel some degree of obligation toward a number of persons who have contributed to his study, sentimental expressions of "deep gratitude" to his advisor, his spouse, his typist, and innumerable other contributors are hardly called for. Certainly, helping graduate students with their research has high priority among the responsibilities the faculty is willing to assume. Professional people do not consider this "beyond the call of duty" or deserving of special thanks.

The Table of Contents. Since there is no index in a thesis or dissertation, the table of contents becomes the only means of locating material within the report. Even more important, it provides the framework around which the report is organized and is, therefore, the base of operations. It should be very carefully done. Certainly, a graduate student should be able to organize his report into the chapter headings, sub-headings, and sub-sub-headings called for by his material. It must be consistent in indentation, capitalization, and so on, and, of course, it must agree with the actual organization of the text of the report.

The Chapter-Title Pages. Usually pages headed by a chapter title carry the line CHAPTER . . . about two inches from the top of the page, followed three spaces below by the title of the chapter in capital letters. The first line of text begins three spaces below the title. The page number should be centered two spaces below the last line of the text.

Headings. Major headings are centered on the line, and capitalized on the major words. They are separated from the preceding and the succeeding line of the text by three spaces. Sub-headings are underscored and indented three, five, or seven spaces. Only the first words and proper nouns are capitalized. They are followed by a period, and the text begins on the same line. All headings should maintain parallel structure.

Margins. Margins of 1¾ inches on the left and 1¼ inches on the right-hand sides of the page are commonly used. The last line of writing should be one inch above the bottom edge of the page. Except on chapter-title pages, the page number should come at the top right hand corner of the page, one

inch from the top edge and 1¼ inches from the right edge. The first line of text should be two or three spaces below the page number.

Pagination. Pages should be numbered consecutively in Arabic numerals from the first page of text to the end of the manuscript (including the appendices). The pages in the introductory sections are numbered ii, iii, iv . . . , one inch from the bottom of the page. Exceptions to this rule include the title page, which is counted but not numbered, and the approval page, which is neither numbered nor counted. All page numbers should stand alone, without periods, hyphens, or dashes.

Bibliography. The bibliography tells the reader the sources of the investigator's information; it is always required in a thesis or dissertation. Generally, it should include only sources that have a direct bearing on the study and should be labeled SELECTED BIBLIOGRAPHY rather than simply BIBLIOGRAPHY. It must include every reference used in the footnotes and others of significance to the study. On the other hand, the bibliography must not appear padded. Generally the bibliography is not annotated, since the review of the literature is actually a form of annotation from the standpoint of the study. If the bibliography is extensive, it can be divided into *books* and *periodicals,* each in (numbered) alphabetical arrangements. The bibliography should be preceded by a fly-sheet bearing the word BIBLIOGRAPHY. The fly-sheet and the first page of the bibliography, like all title pages, are numbered at the bottom of the page. The bibliography must, of course, be accurate and complete, since errors and omissions automatically render it useless. It should also be in good form.

Appendix. Material which, though pertinent to the study, would impede the flow of the report rather than aid in its understanding should ordinarily be placed in the appendices rather than in the main body of the text. Thus, while summary tables and other material necessary for interpretation of the study are placed in the text, tables of raw data should be put in the appendix. Similarly, the discussion of the general nature and orientation of the questionnaire used must be included in the chapter on the design of the study, but the actual questionnaire, with the cover and the follow-up letters, should be placed in an appendix. The appendices should be preceded

by a numbered fly-sheet with the word APPENDIX centered on the page.

Footnote and Bibliographic Form. Credit must be given for material that has been borrowed—whether verbatim or paraphrased—from other writers. This is done through a system of footnotes indicated in the text by superscripts placed immediately following the name of the author or source, at the end of the sentence in which the reference to the borrowed source is made, or at the end of the quotation. Although any of these forms is acceptable, the first is easier to handle, especially in such instances as: "Both Smith[1] and Brown[2] express. . ." Footnotes can be numbered consecutively throughout the paper or consecutively throughout the chapter.

Stylistic details pertaining to footnotes and bibliographic entries vary to such an extent that it is relatively impossible to cover all situations. A number of "acceptable forms" for some of the more common types of entries are illustrated below.

Books

Footnotes: John C. Smith and Robert B. Case, *Principles of Research* (New York: Doe, 1962), p. 111.

Bibliography: SMITH, JOHN C. and CASE, ROBERT B. *Principles of Research.* New York: Doe, 1962.

Later references may be abbreviated *Ibid.,* or *Ibid.,* p. 112 if the second reference follows immediately. Smith and Case, *op. cit.,* or Smith and Case, *op. cit.,* p. 112 is used when the reference is made later in the same chapter.

Periodicals

Footnotes: John C. Smith, "A Study of Common Errors in Spelling," *Journal of English Usage,* 16 (April, 1937): 116–66.

Bibliography: SMITH, JOHN C. "A Study of Common Errors in Spelling," *Journal of English Usage,* 16 (April, 1937): 116–66.

Articles in an Encyclopedia

Footnotes: John C. Smith, "Spelling," in James L. Brown (ed.), *Encyclopedia of Language* (New York: Doe, 1950), pp. 345–356.

Bibliography: SMITH, JOHN C. "Spelling," in Brown, James L. (ed.). *Encyclopedia of Language*. New York: Doe, 1950.

Unpublished Material

Footnotes: John C. Smith, "An Investigation of Common Errors in Scientific Writing" (Unpublished Master's Thesis; New York: University of New York, 1935).

Bibliography: SMITH, JOHN C. "An Investigation of Common Errors in Scientific Writing." Unpublished Master's Thesis; New York: University of New York, 1935.

Another system used by such publications as the *Review of Educational Research* makes the bibliography serve as footnotes. The bibliography is arranged alphabetically and numbered consecutively. Reference to a given entry is made by inserting its number in a parenthesis immediately following the name of the author—for example, "Smith (79:174) reports . . ." where "79" is the number of the Smith entry in the bibliography, and "174" is the page in Smith's book on which the particular reference is to be found. Since articles tend to be short, the specific page number is not generally included in the parenthesis when the reference is to an article. This system is less convenient for the reader and is generally accepted only when there are so many references to be listed—some more than once—that footnotes would take up a substantial part of many of the pages of the text.

Quotations. Wholesale use of quotations is to be discouraged, since most quotable material can be paraphrased (with citation) to better advantage for the particular orientation of the report. Quotations are appropriate: 1. when a point is to be challenged or there is need for special clarity in the issue being engaged—for example, when a point of law is involved; 2. when two conflicting positions are to be compared; and 3. when a point is so well stated, perhaps by a recognized authority, that it would add prestige to the idea being expressed and respectability to the study. It must be remembered, however, that notes should be taken for their significance rather than for their wit or literary flavor. The inclusion of a non-

pertinent quotation simply because it comes from an authority, or is cleverly stated, is an indication that the investigator is not too clear as to what he is about.

Short quotations generally are included as part of the regular text with quotations marks; longer quotations are indented and single-spaced without quotations marks. All quotations are footnoted. Omissions from a quotation are indicated by three spaced dots with an additional dot to represent the period when the omission occurs at the end of the sentence —for example, "The validity of a measuring instrument . . . is frequently difficult to determine. . . ."

Tables. The usual format of a table is shown below.

TABLE 1

Representativeness of the Sample

	Population		Sample	
Stratum	Number	Percent	Number	Percent
1	2803	11.9	67	12.1
2	2759	11.7	63	11.3
3	3550	15.1	87	15.6
4	2638	11.2	63	11.3
. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
9	2119	9.1	51	9.2
Total	23 481	100.0	556	100.0

The word TABLE and the table title are capitalized, the table number is in Arabic numerals. Double lines appear at the top and a single line separates the heading section. The title should be concise; avoid, for instance, "TABLE SHOWING. . . ." Tables should not be complicated; rather than attempting to cover too many points in one table, make two tables. Vertical lines should not be used to separate the columns: to be understandable and attractive, a table should not be so crowded that vertical lines are necessary. Footnotes to tables should be referenced through such symbols as *, †, ‡, §, ||, and ¶ (in order) rather than through superscripts that might tend to be confused with the numbers in the table. Also note that tables should not be broken to fall on two pages, unless they are too long to fit on one. When a table must be placed lengthwise on a page, it should be placed so the title is next to the binding.

TYPING

Typing is always a chore. Generally, if the student can type he probably should type his rough drafts, inasmuch as he gains insight into his study by going through the actual motions of putting it together. His first drafts will have to be revised and he might consider using triple spacing in order to permit easy correction and revision.

A carbon copy is an absolute necessity. Copies of basic data should also be made (by thermofax, if necessary) and one copy stored in a safe place.

Besides being insurance against what might prove extremely costly in the event of loss or misplacement, a carbon copy also permits the student to continue with his work even when the manuscript is in the hands of his advisor or the typist.

Generally, the student should not type his own final draft. Unless he is an expert typist, he probably should allow himself the luxury of having the final step toward his degree custom-made. The final draft calls for many carbons and almost erasure-free copy, a task that is better placed in the hands of a professional typist. On the other hand, the student must assume all responsibility for providing the typist with a usable copy and for checking her work. Obviously, the better the draft presented to the typist, the fewer the errors and the better the arrangement of the final product. A good typist can catch an occasional error, but a sloppy rough draft is an open invitation to unsatisfactory work.

STYLE

Effective writing

Invariably, the greatest weakness of the research report is in expression and organization. Faculty advisors—almost without fail—spend more time on the grammatical and organizational aspects of the report than they do on the research design. Frequently, the trouble is a matter of carelessness—certainly, a graduate student should be able to make verbs agree with subjects, to avoid split infinitives, and to express his ideas clearly and effectively.

Unfortunately, however, inability to write is not restricted to graduate students; educators in general have been charged by a number of writers with similar incompetence. Shannon,[2] for example, points out that "the evidence [of incompetence or indifference in writing] against education is too consistent and too convincing to be shrugged off as inconsequential." He goes on to say that dull writing is no more excusable than a dull wit, that effective writing is an art that can be learned, and that it is hard to believe that a person intelligent enough to do a reputable piece of research cannot report what he did with an equal degree of competence.

Dullness and monotony probably head the list of the specific criticisms of educational writings, but ineffectiveness and incoherence in organization, lack of clarity and forcefulness in sentence and paragraph structure, lack of precision in vocabulary, and even common grammatical errors are also frequently mentioned. Other faults more directly related to the research report include lack of precision in the statement and delineation of the problem, the drawing of sweeping generalizations, the use of flowery and ineffective language, and the inadequate synthesis of the various parts of the report.

To make matters worse, this is an area in which it is difficult to give constructive advice. The advisor is occasionally faced with having to rewrite each sentence and reorganize one section after another, or to give fatherly advice of the variety of "Be more careful," "Write better," or "Rewrite this section"—all of which is rather futile. Another consideration concerns the wisdom of twisting every sentence to the advisor's writing style; each person has his own way of saying things which generally should be respected. Undoubtedly, many research reports need to be improved, but there is a need for some degree of tolerance and appreciation of the fact that there is frequently more than one style of good writing. On the other hand, poor writing and carelessness cannot be encouraged or condoned—and there is place for telling the student to start over again.

Some graduate students have not learned how to organize a research paper. The advisor would do well to insist that, be-

[2] John R. Shannon, "Art in Writing for Educational Periodicals," *Journal of Educational Research*, 44: 599–610, April 1951.

fore he undertakes to write one single line of his report, the student read half-a-dozen of the better organized and better written theses or dissertations in his field. With the small cost of microfilm, providing such models should impose far less financial strain on the institution than the failure to do so imposes in academic stress upon the advisor. Thus, a student familiar with good thesis writing would know that 1. the research report is written in the past tense (for example, the study was conducted . . . , Group A gained . . . , and so on.) On the other hand, the conclusions, supposedly representing a fact applicable to more than the single instance, are written in the present (for example, motivation *is* conducive to effective learning). Certain parts of the description of the locale of the study should also be in the present (for example, Miami is located on the Gold Coast of Florida). 2. Personal pronouns of the first person (I, me, my, we, and our) are to be avoided; a research report should be impersonal. In general, references to the writer or the investigator should be kept to a minimum. 3. The active voice is generally to be preferred to the passive. 4. Numbers up to one hundred and all numbers beginning a sentence should be written out in full. The local thesis manual should be consulted for more specific and detailed suggestions. Specialized sources such as the *Chicago Manual of Style*[3] and the American Psychological Association *Publication Manual*[4] should be consulted for unusual problems that may arise.

Vocabulary

The purpose of the report is to communicate to colleagues what was done and what was found—not to impress them with one's vocabulary or with one's ability to understand something "so obviously complicated." The writing should, therefore, aim for simplicity, clarity, and conciseness; it should avoid attempts at flowery language and other forms of verbal gymnastics. This does not imply the exclusive use of simple vocabulary, but rather the "art of plain talk" within the framework of the complexity of the materials presented.

[3] Kate L. Turabian, *A Manual of Style* (rev. ed.; Chicago: University of Chicago Press, 1949).

[4] American Psychological Association, *Publication Manual* (rev. ed.; Washington, D.C.: The Association, 1957)

Sentence Organization

Every sentence must convey its meaning with the maximum degree of precision. It is a good exercise, for example, to go over the report with a view to removing phrases, sentences, and even paragraphs that contribute nothing to the report. It is not uncommon, for instance, to see a paragraph begin with "Dr. John K. Smith of the University of . . . in 1950 wrote an article entitled: ". . ."; or even, "Smith conducted a study . . . . He found. . ." Why all the verbiage, when the footnote contains all the required information?

The report should be organized to promote the straightest path to reader comprehension. This is sometimes complicated by the fact that the investigator is so familiar with his topic that he loses perspective and cannot see what needs to be said, what needs to be emphasized, and what can be left out. As a result, he may leave out a key idea and leave the poor reader perplexed on a point that is crucial to his understanding. Before giving the report a final polish, therefore, it is generally advisable to have some other person read the rough manuscript with a view to improving its general organization.

THE RESEARCH REPORT

No matter how significant a study may be, it is useless if it is not reported. If we spend the time and energy to conduct a research project, it becomes a professional responsibility to make the results available to others, for this is the only way in which the profession can prosper. Publication also makes it possible to avoid duplication.

The reasons for an investigator not reporting a study are varied. He may feel that the reporting of his study is something of an anti-climax; he may have been with the project so long that it has become stale. Some investigators feel that they are interested only in the results and that the labor and precision required for writing a complete report loom rather large, especially since many find effective writing difficult.

A partial solution that has much to recommend it is for the investigator to write his report as he goes along. The review of the literature and the design of the study should be written in semi-final form before the study is undertaken. For the

investigator to put himself in a position to have to explain what he is doing and why is obviously one of the most effective ways for him to clarify his thinking concerning his problem and the procedures he should follow in its investigation. Not only will it bring his study into focus, it also will result in an improvement both in the design of study and the adequacy of the report. Obviously, the written report cannot be in final form until the whole study is completed. The investigator may have to rewrite his problem in line with the conclusions he has been able to derive from his data, but he should take every opportunity to project his study as far as he can. He should, for example, anticipate the type of data he is likely to obtain, the type of analysis he will be able to make, and even the type of tables he will need to devise. When this is done systematically, the writing of the report is not a chore but a companion step to the investigation itself.

Even when the study is conducted primarily to fulfill a degree requirement, some attempt should be made to publish an article that will bring the study to the attention of the profession. This is especially true of master's theses, which often remain relatively unknown. It is also possible that a doctoral candidate can make a contribution by publishing an article on an aspect of his study which is not adequately covered in his summary in *Dissertation Abstracts.*

This, of course, raises the question of the contributions such articles might make to the advancement of education. Obviously, anything that would advance the cause of education should be made available, and, as a general rule, anything that a member of the profession in good standing has taken time and energy to investigate probably is not devoid of merit—it must contain at least one idea that will stimulate others. The converse viewpoint is that the professional literature is so cluttered with articles—many of poor quality—that journal space should be restricted to articles that make a significant contribution.[5,6]

[5] Frymier reports 788 manuscripts turned down in one year by the editors of seven journals for reasons of faulty design, faulty interpretation of data, triviality of the problem, poor writing, unsuitability with respect to the particular journal, and so on.

[6] Jack R. Frymier, "Problems in Reporting Research," *Phi Delta Kappan*, 40 (June, 1959): 376–7.

Another aspect of publication to be considered is the need for concise writing. In view of the shortage of journal space, it becomes a matter of professional courtesy for the writer to strive for brevity within the framework of the necessary clarity. It may be possible to say in one sentence what might normally be said in two and to tabulate material which would take pages to describe in detail. This, of course, has its drawbacks in that overemphasis on brevity may result in the omission of certain details that would make the study meaningful.

EVALUATION OF THE RESEARCH REPORT

The production of a good research report calls for meticulous attention to innumerable details involving all the points considered here and many more. This is a crucial aspect of graduate training and the quality of the final product is a direct reflection of the "quality" of the student. It must be recognized that quality in as complex a matter as a thesis or dissertation is frequently a matter of intangibles which are difficult to identify. It is generally easier to identify areas of weakness that make a report inadequate than it is to spell out the specific points, attention to which will guarantee its adequacy. It must be realized that producing a good thesis—just like winning a football game—is more than mere compliance with rules.

Judging the adequacy of a research report after it has been submitted—though even then it is a difficult task—is apparently easier than listing definite criteria which will fit every study that might be undertaken. The following items are, therefore, simply some of the points a student might consider in the evaluation—and, especially, in the improvement—of his report.

1. *General format* — attractiveness; conformity to external mechanics of good form: margins, pagination, and so on; freedom from typographical errors and erasures.

2. *Title* — appropriateness to the problem actually investigated; clarity and conciseness.

3. *Problem* — significance and possible contribution; clarity and conciseness of the statement of the problem; parsimony and tenability of the

basic hypotheses; feasibility and suitability of the study.

4. *Review of the Literature* — thoroughness and comprehensiveness; evaluation and synthesis of the sources.

5. *Design* — adequacy and appropriateness to the problem under investigation; adequacy of the description of the design; adequacy of the instruments and procedures.

6. *Analysis of the Data* — validity of the data; reliability; adequacy of the analysis; appropriateness of statistical procedures; objectivity and insight in interpretation; significance of the tables and other means of presentation.

7. *Conclusions* — validity of the conclusions; foundation on basic evidence; recognition of assumptions and limitations; integration with the statement of the problem; synthesis of the status of the problem and suggestions for further investigations.

8. *General Scholarship* — logical and coherent organization; breakdown into an effective system of headings and sub-headings; evidence of insight into the nature of the problem; imagination in the design of the study and the interpretation of the results; evidence of adequate grasp of research and statistical tools; display of a scientific attitude; effectiveness in presentation of the report.

SELECTED REFERENCES

COLTON, F. BARROWS. "Some of My Best Friends are Scientists," *Scientific Monthly,* 69 (September 1949) : 156–60.

DAVIS, ROBERT A. "Writing a Thesis in Education," *Peabody Journal of Education,* 27 (March 1950) : 285–95.

DVORAK, EARL A. "General Guide to a Study of Research Reports," *Peabody Journal of Education,* 34 (November 1956) : 141–4.

ELLISON, WILLIAM H. "And Gladly Teach but Publish We Must," *Educational Forum,* 24 (March 1960) : 321–7.

FLESCH, RUDOLPH F. *The Art of Plain Talk.* New York: Harper, 1946.

———. *The Art of Readable Writing.* New York: Harper, 1949.

———. *How to Make Sense.* New York: Harper, 1954.

FLESCH, RUDOLPH F. and LASS, A. H. *The Way to Write*. 2nd ed.; New York: McGraw-Hill, 1955.

FRYMIER, JACK R. "Problems in Reporting Research," *Phi Delta Kappan,* 40 (June 1959) : 376–7.

GOOD, CARTER V. "Effective Reporting of Research," *Phi Delta Kappan,* 24 (December 1941) : 178–84.

GUNNING, ROBERT. *The Technique of Clear Writing*. New York: McGraw-Hill, 1952.

JOHNSON, GRANVILLE B. "A Method of Evaluating Research Articles in Education." *Journal of Educational Research,* 51 (October 1957) : 149–51.

JOHNSON, LOAS W. "Educational Research and its Dissemination," *Educational Leadership,* 10 (April 1953) : 423–7.

LINDVALL, CARL M. "Improving Research: The Review of Related Research," *Phi Delta Kappan,* 40 (January 1959) : 179–80.

MCCANN, LLOYD E. "Presenting that Idea in the Professional Journal," *Phi Delta Kappan,* 39 (January 1958) : 173–6.

SHANNON, JOHN R. "Style in Reporting Research," *California Journal of Educational Research,* 1 (November 1950) : 215–8.

———. "Tips to Writers from Seventy-Five Editors of Educational Periodicals," *Journal of Educational Research,* 44 (December 1950) : 241–68.

———. "Art in Writing for Educational Periodicals," *Journal of Educational Research,* 44 (April 1951) : 599–610; 46 (January 1953) : 333–45; 47 (March 1954) : 489–504.

STRANG, RUTH M. "Principles of Readability Applied to Reporting Research," *Teachers College Record,* 49 (April 1948) : 449–51.

SYMONDS, PERCIVAL M. "Characteristics of a Good Dissertation in Education," *Teachers College Record,* 24 (January 1933) : 312–26.

———. "Common Faults in Graduate Research in Education," *Journal of Educational Research,* 27 (March 1934) : 481–92.

TURABIAN, KATE L. *Manual for Writers of Term Papers, Theses, and Dissertations*. Chicago: University of Chicago Press, 1955.

Index

EURASIA'S BOOKS ON EDUCATION & PSYCHOLOGY

	Rs P
AN INTRODUCTION TO GUIDANCE *by* Lester D. Crow and Alice Crow	10 00
INTRODUCTION TO EDUCATION *by* Lester D. Crow and Alice Crow	10 00
EDUCATION IN SECONDARY SCHOOL *by* Lester D. Crow and Alice Crow	8 00
NEW DIRECTIONS IN THE AMERICAN UNIVERSTY *by* Dr. John T. Reid	4 50
EVALUATION IN MODERN EDUCATION *by* J. Wayne Wrightston, Joseph Justman & Irving Robbins	8 00
TESTING FOR TEACHERS *by* Henry E. Garrett	6 00
GENERAL PSYCHOLOGY *by* Henry E. Garrett	10 00
BASIC PRINCIPLES OF STUDENT TEACHING *by* Harold P. Adams & Frank G. Dickey	4 00
SOCIAL PSYCHOLOGY *by* Hubert Bonner	8 00
EDUCATIONAL PSYCHOLOGY *by* Lester D. Crow and Alice Crow	9 00
HUMAN DEVELOPMENT & LEARNING *by* Lester D. Crow and Alice Crow	8 00
ELEMENTARY STATISTICS FOR TEACHERS *by* Manuel	In Press
THE SCIENCE OF EDUCATIONAL RESEARCH *by* Mouly	,,
BASIC PRINCIPLES OF SUPERVISION *by* Harold P. Adams and Frank G. Dickey	,,
ELEMENTARY SCHOOL ADMINISTRATION & SUPERVISION *by* Elsbree and Mc. Nally	,,
ELEMENTARY SCHOOL CURRICULUM FROM THEORY TO PRACTICE *by* Jameson and Hickes	,,